

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2006, 19:07:45 ; Search time 6252 Seconds

(without alignments)  
10731.394 Million cell updates/sec

Title: US-10-520-820-28

Perfect score: 1434

Sequence: 1 atgaagaatcgcctgcacga.....aacagataaaaaagcctaa 1434

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

EST:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	820	57.2	820	10	CL659093 PRI0133A
2	489.8	34.2	895	10	CL655529 PRI0149D
3	419.4	29.2	682	9	CC871613 NDL.2A24.
4	323	22.5	722	8	DN468758 USDA-PP 1
5	252	17.6	879	1	B2569854 msh2.105.
6	237.2	16.5	682	6	CB065364 EST645045
7	234.4	16.3	593	10	CL678544 PRI0123B
8	178	12.4	1135	9	B2564465 pac82.164
9	159.6	11.1	879	10	AG841316 Oryza sat
10	159	11.1	762	10	CL690881 PRI0155a
11	146.8	10.2	1195	9	B2554342 pac81.60.
12	100.8	7.0	505	2	BS32786 UB53d03.X
13	96	6.7	226	9	B2648805 OGCAM49TM
14	96	6.7	251	9	B2648800 OGCAM49TM
15	90.4	6.3	764	7	CN817146 HRO4528 G
16	90.4	6.3	692	10	CM332900 104.830.1
17	84	5.9	483	9	BH404405 AG-ND-140
18	73.2	5.1	550	9	BH390803 AG-ND-140
19	66.4	4.6	502	10	AG945142 Drosophila
20	63.8	4.4	1474	9	B2552978 pac81.60
21	54.8	3.8	622	9	CC821631 MBSP55 U
22	54.6	3.8	623	7	COL40047 EST834718

23	54.6	3.8	626	7	COL42807	COL42807 EST837478
24	53	3.7	615	7	COL41185	COL41185 EST835856
25	51	3.6	785	6	CP447053	CP447053 EST683398
26	47.4	3.3	959	9	AF075847	AF075847 AF075847
27	46.4	3.2	878	10	CM942436	CM942436 TCB23.4.H
28	45.8	3.2	502	7	CN898041	CN898041 010701AAZ
29	45.2	3.2	584	5	CA030018	CA030018 HX05N16r
30	45.2	3.2	621	8	DT002298	DT002298 MRC1016
31	44.8	3.1	715	8	DR479547	DR479547 W602817.C
32	44.8	3.1	715	8	DR485191	DR485191 W602817.C
33	44.8	3.1	715	8	DR486049	DR486049 W602817.B2
34	44.8	3.1	723	8	DR480409	DR480409 W602817.BR
35	44.6	3.1	414	11	P957R	AL390661 Leishmani
36	44.6	3.1	776	6	CX545158	CX545158 UCRPT01.5
37	43.6	3.0	652	5	B0996249	B0996249 OCG12R01
38	43.6	3.0	728	8	DR667881	DR667881 ZM_BBP008
39	43.6	3.0	736	5	BQ850159	BQ850159 QGB1107
40	43.6	3.0	772	10	CZ002297	CZ002297 OA_BBA005
41	43.4	3.0	634	7	CN910981	CN910981 021119ABM
42	43.2	3.0	453	3	BJ935968	BJ935968 BJ935968
43	42.6	3.0	380	7	CO338482	CO338482 EN16913.5
44	42.6	3.0	430	7	CO292447	CO292447 EK200212
45	42.6	3.0	616	3	BI639334	BI639334 SD21893.5

## ALIGNMENTS

RESULT 1  
LOCUS CL659093 820 bp DNA linear GSS 09-JUL-2004  
DEFINITION PRI0133A\_G12 - PRI0133A.B21 (820) Mixed stage fosmid library of P. pacificus var. California Pristionchus pacificus genomic, genomic survey sequence.

ACCESSION CL659093  
VERSION CL659093.1 GI:50142804  
KEYWORDS GSS.  
SOURCE Pristionchus pacificus  
ORGANISM Pristionchus pacificus  
Bukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida; Neodiplogasteridae; Pristionchus.

REFERENCE 1 (bases 1 to 820)  
Srinivasan, J., Otto, G.W., Kahlow, U., Geisler, R. and Sommer, R.J. 1999. A database for the nematode Pristionchus pacificus. This library was generated at Caltech, Pasadena, USA and end sequenced at Vancouver, Canada.  
Nucleic Acids Res. 32 (1), D421-D422 (2004)

JOURNAL PUBMED  
COMMENT Contact: Sommer RJ  
Evolutionary Biology  
Max-Planck-Institute for Developmental Biology  
Spemannstr. 37-39, Tuebingen D-72076, Germany  
Tel: 00497071601371  
Fax: 00497071601498  
Email: ralf.sommer@tuebingen.mpg.de  
This library was generated at Caltech, Pasadena, USA and end sequenced at Vancouver, Canada.  
Seq primer: T7  
Class: fosmid ends.  
Location/Qualifiers  
1. 820  
/organism="Pristionchus pacificus"  
/mol\_type="genomic DNA"  
/strain="California"  
/db\_xref="taxon:54126"  
/clone\_lib="Mixed stage fosmid library of P. pacificus var. California"  
/note="Vector: pBplfos-5 Fosmid vector"

## ORIGIN

Query Match 57.2%; Score 820; DB 10; Length 820;  
Best local similarity 100.0%; Pred. No. 3.6e-225;  
Matches 820; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 214 TTGACGGGCAATGACGATGACGCGCGCTGAGTAATCTCTGCGGACGTC 273  
 DB 1 TTGACGGGCAATGACGATGACGCGCGCTGAGTAATCTCTGCGGACGTC 60  
 QY 274 AAATGCGAATCTGTTTCTGTAACGACGATCCGACATACCAATTAACGGGTA 333  
 DB 61 AAATGCGAATCTGTTTCTGTAACGACGATCCGACATACCAATTAACGGGTA 120  
 QY 334 CGGAAACCAACGATGATCCGCTGAGTATTTGAAAGAGTTTCGAAGGTTGATCCG 393  
 DB 121 CGGAAACCAACGATGATCCGCTGAGTATTTGAAAGAGTTTCGAAGGTTGATCCG 180  
 QY 394 CCGCTGACGACGAGGATTAATCAGGCGCTGAGTTCGATGCGCGCTGCTTTCGAC 453  
 DB 181 CCGCTGACGACGAGGATTAATCAGGCGCTGAGTTCGATGCGCGCTGCTTTCGAC 240  
 QY 454 TACGCGCAAGGTGCGCTGCGCAAGGATGATGATCCGACGCGCTGAAAGCGG 513  
 DB 241 TACGCGCAAGGTGCGCTGCGCAAGGATGATGATCCGACGCGCTGAAAGCGG 300  
 QY 514 GTTCGCGGCTGATGATCCGAAAGGATACGATTTTGAGCGCTACGCGCGCTACG 573  
 DB 301 GTTCGCGGCTGATGATCCGAAAGGATACGATTTTGAGCGCTACGCGCGCTACG 360  
 QY 574 TTAACGCGCAATCTCTCGGAATTTGAAGCTGTTGCTGTAATGTAAGACCGAAG 633  
 DB 361 TTAACGCGCAATCTCTCGGAATTTGAAGCTGTTGCTGTAATGTAAGACCGAAG 420  
 QY 634 ATTGTTGAGCGCGGATGAAATCTGATTCGATTAACGAACTCTGCTGTTTGTG 693  
 DB 421 ATTGTTGAGCGCGGATGAAATCTGATTCGATTAACGAACTCTGCTGTTTGTG 480  
 QY 694 CGTTCGCAACGAGGATGATGCTGCTGCAACCGGCTGAAAGCGCGCTGATATCC 753  
 DB 481 CGTTCGCAACGAGGATGATGCTGCTGCAACCGGCTGAAAGCGCGCTGATATCC 540  
 QY 754 CAAACGCGAAGATGTAATGACGTTACCGGTGCGGCGCAACGATGATGCGCTG 813  
 DB 541 CAAACGCGAAGATGTAATGACGTTACCGGTGCGGCGCAACGATGATGCGCTG 600  
 QY 814 GCAACGCTGCGAGCGGCTAATTCGCTGGAAGAACCTGCTTCTTTCGCAATGCG 873  
 DB 601 GCAACGCTGCGAGCGGCTAATTCGCTGGAAGAACCTGCTTCTTTCGCAATGCG 660  
 QY 874 GCGCTGCTGCTGCGCAACTGGAACCTCAACGCTTTCGCGATGAGTGAATAATG 933  
 DB 661 GCGCTGCTGCTGCGCAACTGGAACCTCAACGCTTTCGCGATGAGTGAATAATG 720  
 QY 934 GTACGTGACGCTGACGATGACAGGCTTTCGCTGATGACCGAAGAGAACTGA 993  
 DB 721 GTACGTGACGCTGACGATGACAGGCTTTCGCTGATGACCGAAGAGAACTGA 780  
 QY 994 GTACGCGCAAGCGCTGAAACCTGCTGAAAGAAAGTGTGATGA 1033  
 DB 781 GTACGCGCAAGCGCTGAAACCTGCTGAAAGAAAGTGTGATGA 820

RESULT 2  
 CL665529/c 895 bp DNA linear GSS 09-JUL-2004  
 LOCUS PR1019d.H06 - PR1019d.B21 (895) Mixed stage fosmid library of P.  
 DEFINITION pacificus var. California Pristionchus pacificus genomic, genomic  
 survey sequence.  
 ACCESSION CL665529  
 VERSION CL665529.1 GI:50157226  
 KEYWORDS GSS.  
 SOURCE Pristionchus pacificus  
 ORGANISM Pristionchus pacificus  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
 Nematoda; Pristionchidae; Pristionchus.  
 REFERENCE 1 (bases 1 to 895)  
 AUTHORS Strinivasan,U., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.U.  
 TITLE Appad: an Acedb database for the nematode satellite organism

JOURNAL Pristionchus pacificus  
 PUBMED Nucleic Acids Res. 32 (1), D421-D422 (2004)  
 14681447  
 COMMENT Contact: Sommer RJ  
 Evolutionary Biology  
 Max-Planck-Institute for Developmental Biology  
 Spemannstr. 37-39, Tuebingen D-72076, Germany  
 Tel: 00497071601371  
 Fax: 00497071601498  
 Email: ralf.sommer@uebingen.mpg.de  
 This library was generated at Caltech, Pasadena, USA and end  
 sequenced at Vancouver, Canada.  
 Seq primer: T7  
 Class: fosmid ends.  
 Location/Qualifiers  
 1..895  
 /organism="Pristionchus pacificus"  
 /mol\_type="genomic DNA"  
 /strain="California"  
 /db\_xref="taxon:54126"  
 /clone\_lib="Mixed stage fosmid library of P. pacificus  
 var. California"  
 /note="Vector: pBf108-5 Fosmid vector"

ORIGIN  
 Query Match 34.2%; Score 489.8; DB 10; Length 895;  
 Best Local Similarity 97.5%; Pred. No. 8,7e-130;  
 Matches 508; Conservative 0; Mismatches 12; Indels 1; Gaps 1;

QY 915 GATGAGCTGGAATAATGCTGTGATGATGATGATGATGATGATGATGATGATG 974  
 DB 895 GATGAGCTGGAATAATGCTGTGATGATGATGATGATGATGATGATGATGATG 836  
 QY 975 AGAGAACTGAAGCTGAGCGGCTGAGCGGCTGAAAGCT-AGTGAATAAGTGTATGA 1033  
 DB 835 AGAGAACTGAAGCTGAGCGGCTGAGCGGCTGAAAGCT-AGTGAATAAGTGTATGA 776  
 QY 1034 CCAACGCTGCTTTCGATCTGCAACGCGGCAAGCTCTTATCTGCAATATGCGCA 1093  
 DB 775 CCAACGCTGCTTTCGATCTGCAACGCGGCAAGCTCTTATCTGCAATATGCGCA 716  
 QY 1094 AGCTGGGTGACCGCTGATGATGATGATGATGATGATGATGATGATGATGATG 1153  
 DB 715 AGCTGGGTGACCGCTGATGATGATGATGATGATGATGATGATGATGATGATG 656  
 QY 1154 GGGATTCGCGCGCGGTAACCCGATGCAAGCGTATGATGATGATGATGATGATG 1213  
 DB 655 GGGATTCGCGCGCGGTAACCCGATGCAAGCGTATGATGATGATGATGATGATG 596  
 QY 1214 CCGTCACTGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1273  
 DB 595 CCGTCACTGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 536  
 QY 1274 TGCAGATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1333  
 DB 535 TGCAGATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 476  
 QY 1334 AAGTCTGGGCAACGAGTGGGCAAGTGTGATGATGATGATGATGATGATGATG 1393  
 DB 475 AAGTCTGGGCAACGAGTGGGCAAGTGTGATGATGATGATGATGATGATGATG 416  
 QY 1394 CCAACATCATCAAGAGATCAACAGAGATCAAGAGATCAAGAGATCAAGAGAT 1434  
 DB 415 CCAACATCATCAAGAGATCAACAGAGATCAAGAGATCAAGAGATCAAGAGAT 375

RESULT 3  
 CC871613 682 bp DNA linear GSS 24-JUL-2003  
 LOCUS ND1.2A24.Spe Notre Dame Liverpool Aedes aegypti genomic clone  
 DEFINITION Notre Dame Liverpool-2A24, genomic survey sequence.  
 ACCESSION CC871613  
 VERSION CC871613.1 GI:33231623

KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

G98.  
Aedes aegypti (yellow fever mosquito)  
Aedes aegypti  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;  
Culicidae; Culicini; Culicini; Aedes; Stegomyia.  
1 (bases 1 to 682)  
Loftus, B., Shetty, J., Knudson, D. and Severson, D.  
BAC end sequencing of Aedes aegypti  
Unpublished (2003)  
Other GS88: ND.2A24.T7  
Contact: Brendan Loftus  
Department of Eukaryotic Genomics  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-3543  
Fax: 301-838-0208  
Email: ente@tigr.org  
Library was provided by David Severson  
Seq primer: SP6  
Class: BAC ends.

FEATURES  
Source  
Location/Qualifiers  
1..682  
/organism="Aedes aegypti"  
/mol\_type="genomic DNA"  
/strain="Liverpool"  
/db\_xref="taxon:7159"  
/clone="Notredame Liverpool-2A24"  
/clone\_lib="Notre Dame Liverpool"  
/note="Vector: pECBAC1, Site 1: Hind III; The library was  
prepared from whole body tissue of newly hatched L1 larvae  
by David Severson at the University of Notre Dame and  
Hongbin Zhang"

ORIGIN  
Query Match 29.2%; Score 419.4; DB 9; Length 682;  
Best Local Similarity 76.2%; Pred. No. 1.8e-109;  
Matches 516; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

QY 286 GTTTCGTGACGAGCATCCGACATTCACCAATTAACGGGTACTTCCCGACCAACAG 345  
Db 1 GTCTCGGTGCGGACCCATCCGACCATCAACAGCTGCGGTGCTGCGGACCAACAG 60

QY 346 CTGATCCGTCTGATTTTGAAGAGCTTTCGAAGTGTATTCGCGACCGCTGACAG 405  
Db 61 CTGATCCGTCTGATTTTGAAGAGCTTTCGAAGTGTATTCGCGACCGCTGACAG 120

QY 406 CGGATTAATCAGGCGCTGATTTGAGAGCTTTCGAGCTTTCGACTACGCCAAGT 465  
Db 121 CGCATTCAGAGGCGCTGATTTGAGAGCTTTCGAGCTTTCGACTACGCCAAGT 180

QY 466 GCGCTGCGAAGCGTACAGCAGATGATCCAACTGCGCGGTAAAGCGGTGTTCCGCTG 525  
Db 181 GCGCTGCGAAGCGTACAGCAGATGATCCAACTGCGCGGTAAAGCGGTGTTCCGCTG 240

QY 526 ATTATATCAAAAGTATCGATTTTGAAGAGCTTTCGAGCTTTCGACTACGCCAAGT 585  
Db 241 ATCATCTCAAAAGTATCGATTTTGAAGAGCTTTCGAGCTTTCGACTACGCCAAGT 300

QY 586 CTCTCGGAATTGAAGCTGTGTGCGTAAATGTAAGACCGAAGAAAGATGTTGAGCGC 645  
Db 301 CTCTCGGAATTGAAGCTGTGTGCGTAAATGTAAGACCGAAGAAAGATGTTGAGCGC 360

QY 646 GCGATGAAGCTGATTCGATTCGAACTCTGCGCTCTGTTAGTACCCGTTCCGAACAG 705  
Db 361 GCGATGAAGCTGATTCGATTCGAACTCTGCGCTCTGTTAGTACCCGTTCCGAACAG 420

QY 706 GGTATGTCGCTGCTGCAACCGGCTAAAGCGCGCTGATTCGAACTCTGTTAGTACCCGTTCCGAACAG 765  
Db 421 GGTATGTCGCTGCTGCAACCGGCTAAAGCGCGCTGATTCGAACTCTGTTAGTACCCGTTCCGAACAG 480

QY 766 GTGATGACGTTTACCGGTGCGGCGAACAAGTGAATGCGCTCTGCGCGCAACCGCTGCA 825  
Db 766 GTGATGACGTTTACCGGTGCGGCGAACAAGTGAATGCGCTCTGCGCGCAACCGCTGCA 825

Db 481 GTGTTGATGATGACCGGCGCGGCGACACCTGATCGGCGTGTGCGCGCTGCTGCGCC 540

QY 826 GCGGTAATTCCTGCGAAGAGCTGCTTTTCCCAATGCGCGGCTGCGCTGCTGCTG 885  
Db 541 GCGGTAATTCCTGCGAAGAGCTGCTTTTCCCAATGCGCGGCTGCGCTGCTGCTG 600

QY 886 GCGAATCGGGAACCTCCACGTTTCCGCGATGAGCTGGAATGCTGATACGTGACGT 945  
Db 601 GCGAATCGGGAACCTCCACGTTTCCGCGATGAGCTGGAATGCTGATACGTGACGT 660

QY 946 GCAGATACAGGCTTTTG 962  
Db 661 GCGAATCGGGAACCTTTTG 677

RESULT 4  
DN468758/c 722 bp mRNA linear EST 01-AUG-2005  
LOCUS  
DEFINITION  
Diaphorina citri CDNA clone WHDc040\_E04 5', mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

DN468758.1 GI:71538215  
EST.  
Diaphorina citri  
Diaphorina citri  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha; Psylliformes;  
Psyllidae; Psyllidae; Diaphorina.  
1 (bases 1 to 722)  
Hunter, W.B., Dang, P.M. and McKenzie, C.L.  
Gene expression in adult Asian citrus psylla, Diaphorina citri  
Kuwayama (Hemiptera: Psyllidae)  
Unpublished (2006)  
Contact: Wayne Hunter, Phat Dang, USDA, ARS  
U.S. Horticultural Research Lab, Subtropical Insect Research Unit  
USDA-ARS  
2001 South Rock Road, Ft. Pierce, FL 34945, USA  
Tel: (772) 462-5988, (772) 462-5940  
Fax: (772) 462-5986  
Email: Whunter@usnrl.ars.usda.gov  
Seq primer: T3 primer.

FEATURES  
Source  
Location/Qualifiers  
1..722  
/organism="Diaphorina citri"  
/mol\_type="mRNA"  
/db\_xref="taxon:121845"  
/clone="WHDc040\_E04"  
/sex="Mixed genders"  
/tissue\_type="whole body"  
/dev\_stage="adults"  
/lab\_host="X11-Blue"  
/clone\_lib="Diaphorina citri Kuwayama (Hemiptera:  
Psyllidae)"  
/note="Vector: pBluescript II SK+, Site 1: SmaI; Site 2:  
XhoI; Source: Colonies maintained by Dr. McKenzie, USHRL,  
Fort Pierce, FL. A high quality EST with at least 200  
contiguous bases at Trace Turner score of 20 or better.  
Library construction by Laura Hunnicutt, USDA, ARS, U.S.  
Horticultural Res. Lab, Ft. Pierce, FL, USA."

ORIGIN  
Query Match 22.5%; Score 323; DB 8; Length 722;  
Best Local Similarity 100.0%; Pred. No. 1.2e-81;  
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 304 CCGACATTAACCAATTAACCGGTACTTCCCGAACAACAGCTGATCCGCTGATTTT 363  
Db 722 CCGACATTAACCAATTAACCGGTACTTCCCGAACAACAGCTGATCCGCTGATTTT 663

QY 364 GAAGAAAGTTTGAAGGTTGATTCGACAGCGCTGCAAGAGGATTAATCAGGCGCTG 423  
Db 662 GAAGAAAGTTTGAAGGTTGATTCGACAGCGCTGCAAGAGGATTAATCAGGCGCTG 603

QY 424 AGTTCGATTCGCGCTGCTGCTTCTGATCCGCAAGGTGCGCTGCGAAGCGTACAG 483  
 DB 602 AGTTCGATTCGCGCTGCTGCTTCTGATCCGCAAGGTGCGCTGCGAAGGTACAG 543  
 QY 484 CAGATGATCCAGCTGCGCGCTGAAGCGGGGTTCGGTCTGATTTATCCAAAAGTACC 543  
 DB 542 CAGATGATCCAGCTGCGCGCTGAAGCGGGGTTCGGTCTGATTTATCCAAAAGTACC 483  
 QY 544 GATTTAGAGCGCTACCGCGCGCTACGCTGTTAAGCGCAATCTCTGGAAATTTGAAGCT 603  
 DB 482 GATTTAGAGCGCTACCGCGCGCTACGCTGTTAAGCGCAATCTCTGGAAATTTGAAGCT 423  
 QY 604 GTTGTGCGTAAATGTAAGCCGA 626  
 DB 422 GTTGTGCGTAAATGTAAGCCGA 400  
 RESULT 5  
 BZ569854 879 bp DNA linear GSS 17-DEC-2002  
 LOCUS meh2.105.x2.meh Pseudomonas aeruginosa genomic clone meh2.105,  
 DEFINITION genomic survey sequence.  
 ACCESSION BZ569854  
 VERSION BZ569854.1 GI:27204915  
 KEYWORDS GSS.  
 SOURCE Pseudomonas aeruginosa  
 ORGANISM Pseudomonas aeruginosa  
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 Pseudomonadaceae; Pseudomonas.  
 REFERENCE 1 (bases 1 to 879)  
 AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,  
 Burns,J.L., Kaul,R. and Olsen,M.V.  
 TITLE Whole-Genome-Sequence variation among multiple isolates of  
 JOURNAL Pseudomonas aeruginosa library  
 COMMENT J. Bacteriol. (2002) In press  
 CONTACT: Chris K. Raymond  
 Genome Center  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA  
 Tel: 2062216954  
 Fax: 2066857244  
 Email: ckraymond@u.washington.edu  
 Class: shotgun.  
 FEATURES  
 source Location/Qualifiers  
 1..879  
 /organism="Pseudomonas aeruginosa"  
 /mol\_type="genomic DNA"  
 /strain="MSH"  
 /db\_xref="taxon:287"  
 /clone="meh2.105"  
 /clone\_1lb="meh"  
 /note="Environmental isolate. Whole genomic shotgun  
 library."  
 ORIGIN  
 Query Match 17.6%; Score 252; DB 9; Length 879;  
 Best Local Similarity 59.5%; Pred. No. 4.2e-61;  
 Matches 454; Conservative 0; Mismatches 307; Indels 2; Gaps 2;

DB 271 GCCATGACCAAGCCGCCCTGACCTGCGCGCGCGCGGAGAAAGTTCGACGTACCG 330  
 QY 782 GTCGCGGCGCACACGATGATTCGCTGCTGCGGCAACGCTGCGAGCGGGATATTCGCTGG 841  
 DB 331 GTCGCGGCGCACACGATGATTCGCTGCTGCGGCAACGCTGCGAGCGGGATATTCGCTGG 330  
 QY 842 AAGAACCTGCTCTTTCGCAATGCGCGCGCTGCGCTGCTGCGCAACCTGGAACCT 901  
 DB 391 CTTCCGCGGCGCTGCGGCAACCTGCGCGCGCTGCGCTGCGCAACCTGGAACCT 450  
 QY 902 CCAACGCTTTCGCGGATGACGTGGAATAAGCTGTACGTGACGTGACATACAGCTTGG 951  
 DB 451 CGCGCATGACGCGCGCGCAACCTGCGCGCGCTGACGCGGACGAGGATTCGACGCTG 510  
 QY 962 GCGTGAATGACCGAAGAGAACTGAAGCTGCGCGCTGACGCGGCGCGTAAACGTGTGA 1021  
 DB 511 GCGTGTGCGGCTGCGAGCAATTCCTGCAATCGAAGACGCGCGCGCGCGGAG 570  
 QY 1022 AAGTGTGATGACCAACGCTGCTTTCGATCCTGACGCGCGGCAAGCTCTTATCTGG 1081  
 DB 571 AAGTGTGCTTACCAATGAGGTGCTTTCGATCCTTACGCGCGGCGCAAGTACCTTAC 630  
 QY 1082 CAATGCGCGCAAGCTGCGGATGACGCTTTCGATTCCTGCAACGAGATGCTTCAACCA 1141  
 DB 631 AACAGCGCGCGCGCGCGCAAGGACCGCTGATGCGAGTCAACGACGCTTGTGCA 690  
 QY 1142 AACGCGTGAAGAGGAGATTCGCGCGCGGTAACCACTGCAACGCGTATGCTGCTGG 1201  
 DB 691 CTGCGCTGAAGAGCGCTTTCGCGCGCGCAT- AACTCGGTGACCGCGCTGCGGCTACTCG 749  
 QY 1202 GCGCACTGGAAGCGGTGACGCTGAGTGTGCTTTGAAGAGACACGCGCGCGCTTGA 1261  
 DB 750 CCGGCGCT-GNNNGGTGAGCTGAGTGTGAGTTCGCCCAAAACATCTCGAAGCGCTTG 808  
 QY 1262 TCCCGCGGATCTTCCAGATCTGCTGCTGTAAGAGCGCGCGCACTA 1304  
 DB 809 TTGAGCAGGTGCGCTCGAGCGTGTGTCAGGCGCGCGCAATA 851  
 RESULT 6  
 CB065364 682 bp mRNA linear EST 21-JAN-2003  
 LOCUS EST645045 HOGA Medicago truncatula cDNA clone HOGA-211, mRNA  
 DEFINITION sequence.  
 ACCESSION CB065364  
 VERSION CB065364.1 GI:27810942  
 KEYWORDS EST.  
 SOURCE Medicago truncatula (barrel medic)  
 ORGANISM Medicago truncatula  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;  
 Medicago.  
 REFERENCE 1 (bases 1 to 682)  
 AUTHORS Hahn,M.G., Ojane-Reuhs,T., Samao,D., Town,C.D., Van Aken,S.,  
 Uteback,T., Cho,J. and Fraser,C.M.  
 TITLE ESTs from roots of Medicago truncatula treated with  
 JOURNAL oligogalacturonides of DP 6-20  
 COMMENT Unpublished (2001)  
 CONTACT: Michael G. Hahn  
 Complex Carbohydrate Research Center  
 University of Georgia  
 220 Riverbend Road, Athens, GA 30602-4712, USA  
 Tel: 706-542-4457  
 Fax: 706-542-4412  
 Email: hahn@ccrc.uga.edu  
 TIGR sequence name: MTMB497K  
 More information is available at: www.medicago.org  
 Seq primer: Skmod (CTA GAA CTA gtc gat CC).  
 FEATURES  
 source Location/Qualifiers  
 1..682  
 /organism="Medicago truncatula"  
 /mol\_type="mRNA"



/cultiVar="A17"  
/db\_xref="taxon:3880"  
/clone="HOGA-211"  
/clone\_type="3 day old seedling roots"  
/dev\_stage="24 hours after treatment in the dark at 26 C  
with 0.5 mg/ml oligogalacturonide (DP 6-20) in the  
presence of 100 ug/ml Gentamicin"  
/lab\_host="X10LR"  
/clone\_1lb="HOGA"  
/note="Vector: plasmid SK-; Site\_1: EcoRI; Site\_2:  
XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA  
was directionally ligated into the Unizap XR vector from  
Stratagene and packaged using Gigapack III Gold packaging  
extracts. Plasmids containing cDNA inserts were excised  
from the recombinant lambda-zap phage using Ex-aseIst  
helper phage and propagated in SOLR cells."

## ORIGIN

Query Match 16.5%; Score 237.2; DB 6; Length 682;  
Best Local Similarity 59.5%; Pred. No. 7.4e-57;  
Matches 401; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

Qy 577 ACCGCGAATCTCTCGAATTGAAAGCTGTTGCTGTAATGTAAGACCGAAGAGATT 636  
Db 2 ACTCGAAGCTCAGCGAGTTCGAAACCATGCGTGGTGGCGTGAAGAAAGATTG 61

Qy 637 GTTGAGCGCGCATGAAATGATTCGATTCGAACTCTCGCTCTGTTAGTACCCGT 696  
Db 62 GTCCCGAAGGCGCTTCAACTGCTGCGAAGTCTCACTGCGCGCATGCTGTAACCCG 121

Qy 697 TCCGAACAGGGTATGCTGCTGCTGCAACCGGGTAAAGCGCGCTGATGCAACCCAA 756  
Db 122 GCGCAATGAGCATGACCTGCTGCTGCAACCGGTGCAACCGCATGCACTGCGCGCG 181

Qy 757 GCGCAGAGATGTAATGATGCTGCTGCGGCGCAACGATGATGCTGCTGCGCGCA 816  
Db 182 GCGCGTAAAGTTCGATGTCACCGGTGCGGCGATACCGTCACTGCACTGCGCGGTG 241

Qy 817 ACGTGGCAGCGGTAATGCTGCTGAAAGAAAGCTGCTCTTGGCAATGCGCGCGTGGC 876  
Db 242 GCCATGCGCTGCGGCGAGCACTGCGCGCATGCTGCGCGCACTGCGCGCGCA 301

Qy 877 GTGGTGTGCGCAAACTGGGAACCTCCACGGTTTGGCGATTCGAGCTGGAATAATGCTGA 936  
Db 302 ATGCTGATCGCAAACTGGGATACCGCGCATGACCGCGCGAGTTGCGCGCGCATC 361

Qy 937 CGTGAAGTCAAGATACAGGCTTTGCGTGAATGACCAAGAGAACTGAAGTGGCGTA 996  
Db 362 CAGCGCAAGAGGAGCTCGAGCGCGGGGTGCTGGGTCTGAGAGCAATGCTGCGCATC 421

Qy 997 GCGGCAAGCGGTAAAGTGTGAAAGAGTGTATGACCAAGCTGCTTTGACATCTTG 1056  
Db 422 GACATATCGCGCGCGCAACAGAGATCTGTTCAACATGCTTCTGACATCTTC 481

Qy 1057 CAGCGCGGCGCACTCTTATCTGCGCAAAATGCCGCAAGCTGGGTGACGCTGATGTT 1116  
Db 482 CAGCGCGGCGCACTGACCTATCTGAGCAAGCTGCTGCCAGGAGATCGCTGATGTC 541

Qy 1117 GCGCTCAACAGCGATGCTCCACCAACGAGTGAAGAGGATTCGCCCGCGTAAACCA 1176  
Db 542 GCGGTCAACAGCGATGCTCCGTCAGCGCGCTCAAGGGGCGAGTCTCGATCAACAT 601

Qy 1177 CTGCAACAGGTATGATTTGCTGGCGCACTGGAAGCGTGAAGTGTGCTTT 1236  
Db 602 GTGATGCGCGCATGCGCTGCTGCTGCTGAGAGCGGTGATCATCACTTC 661

Qy 1237 GAAGAGACAGCGC 1250  
Db 662 CCGAGAGACACCC 675

RESULT 7  
CL678544/c

LOCUS CL678544 593 bp DNA linear GSS 09-JUN-2004  
DEFINITION PRI0123b\_A02\_2 - PRI0123b.BR (593) Mixed stage fosmid library of P.  
pacificus var. California Pristionchus pacificus genomic, genomic  
survey sequence.  
ACCESSION CL678544  
VERSION CL678544.1 GI:50184914  
KEYWORDS GSS.  
SOURCE Pristionchus pacificus  
ORGANISM Pristionchus pacificus  
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
Neodiplogasteridae; Pristionchus.  
1 (bases 1 to 593)  
Srinivasan, J., Otto, G.W., Kahlow, U., Geisler, R. and Sommer, R.J.  
ApplB: an AceB database for the nematode satellite organism  
Pristionchus pacificus  
Nucleic Acids Res. 32 (1), D421-D422 (2004)  
JOURNAL 14681447  
PUBMED  
COMMENT Contact: Sommer RJ  
Evolutionary Biology  
Max-Planck-Institute for Developmental Biology  
Spemannstr. 37-39, Tuebingen D-72076, Germany  
Tel: 00497071601371  
Fax: 00497071601498  
Email: ralf.sommer@uebingen.mpg.de  
This library was generated at Caltech, Pasadena, USA and end  
sequenced at Vancouver, Canada.  
Seq primer: T7  
Class: fosmid ends.

## FEATURES

## source

1. 593  
/organism="Pristionchus pacificus"  
/mol\_type="genomic DNA"  
/strain="California"  
/db\_xref="taxon:54126"  
/clone\_1lb="Mixed stage fosmid library of P. pacificus  
var. California"  
/note="Vector: pBp1fos-5 Fosmid vector"

## ORIGIN

Query Match 16.3%; Score 234.4; DB 10; Length 593;  
Best Local Similarity 99.2%; Pred. No. 4.6e-56;  
Matches 246; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 1188 TATGATTTGCTTGGCGCATCGGAGCGGTGCACTGAGTGGTGTGTTGAAGAGACA 1246  
Db 593 TATGATTTGCTTGGCGCATCGGAGCGGTGCACTGAGTGGTGTGTTGAAGAGACA 534

Qy 1247 CGCGGAGCGCTTATGCTGCGCGGATCTTGCAGATCTGCTGTAAGAGCGCGCACTATA 1306  
Db 533 CGCGGAGCGCTTATGCTGCGCGGATCTTGCAGATCTGCTGTAAGAGCGCGCACTATA 474

Qy 1307 AACCAAGAGATTCGCGGAGTAAAGATCTGGGCCAACGCTGCGAGTGTGTGC 1366  
Db 473 AACCAAGAGATTCGCGGAGTAAAGATCTGGGCCAACGCTGCGAGTGTGTGC 414

Qy 1367 TCACTTTGAAGACGTTGCTGACGACCAACATCATCAAGAGATCCACAGATATAA 1426  
Db 413 TCACTTTGAAGACGTTGCTGACGACCAACATCATCAAGAGATCCACAGATATAA 354

Qy 1427 AAGCTTA 1434  
Db 353 AAGCTTA 346

RESULT 8  
B2564465/c  
LOCUS B2564465 1135 bp DNA linear GSS 17-DEC-2002  
DEFINITION pac82-164\_4762.y2 pac82-164 Pseudomonas aeruginosa genomic clone  
pac82-164\_4762, genomic survey sequence.  
ACCESSION B2564465  
VERSION B2564465.1 GI:27190368  
KEYWORDS GSS.  
SOURCE Pseudomonas aeruginosa



QY 836 CGCTGAGAGAGCCGCTTCTTTCGCAATGCGGCGGCTGCTGCTGGCAAACTCG 895  
 DB 479 CGCTGTGAGAGCCGCTGACATGCGCAACCGCGCGGCGCATGCTGCTGGCAAGCTGG 538  
 QY 896 GAACCTCCACGCTTTCG 913  
 DB 539 GCACCGCACCGCTCAACC 556  
 RESULT 10  
 LOCUS CL690881  
 DEFINITION PRI0155A\_A04.2 - PRI0155A\_BR (762) Mixed stage fosmid library of *P. pacificus* var. California *Pristionchus pacificus* genomic, genomic survey sequence.  
 ACCESSION CL690881  
 VERSION CL690881  
 KEYWORDS GI:50212789  
 SOURCE *Pristionchus pacificus*  
 ORGANISM *Pristionchus pacificus*  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida; Nematodiplogasteridae; *Pristionchus*.  
 1 (bases 1 to 762)  
 Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.  
 AppaB: an Acada database for the nematode satellite organism *Pristionchus pacificus*  
 Nucleic Acids Res. 32 (1), D421-D422 (2004)  
 JOURNAL 1468147  
 PUBMED  
 COMMENT Contact: Sommer RJ  
 Evolutionary Biology  
 Max-Planck-Institute for Developmental Biology  
 Spemannstr. 37-39, Tuebingen D-72076, Germany  
 Tel.: 00497071601371  
 Fax: 00497071601498  
 Email: ralf.sommer@uebingen.mpg.de  
 This library was generated at Caltech, Pasadena, USA and end sequenced at Vancouver, Canada.  
 Seq primer: T7  
 Class: fosmid ends.  
 FEATURES  
 source  
 Location/Qualifiers  
 1..762  
 /organism="Pristionchus pacificus"  
 /mol\_type="genomic DNA"  
 /strain="California"  
 /db\_xref="taxon:54126"  
 /clone\_lib="Mixed stage fosmid library of *P. pacificus* var. California"  
 /note="Vector: pBplfos-5 Fosmid vector"  
 ORIGIN  
 Query Match 11.1%; Score 159; DB 10; Length 762;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-34;  
 Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1276 CCAATCTGCTGCGGAGAAAGCGGCGACATTAACCAAGAGATTCGCGGAGTAAAGAA 1335  
 DB 1 CCAATCTGCTGCGGAGAAAGCGGCGACATTAACCAAGAGATTCGCGGAGTAAAGAA 60  
 QY 1336 GTCTGGCCCAACGCTGGCGAAGTGTGGTCTCAACTTGAAGACGCTGTCGACGACC 1395  
 DB 61 GTCTGGCCCAACGCTGGCGAAGTGTGGTCTCAACTTGAAGACGCTGTCGACGACC 120  
 QY 1396 AACATCATCAAGAGATCCACAGAGATTAAGAGCTTA 1434  
 DB 121 AACATCATCAAGAGATCCACAGAGATTAAGAGCTTA 159  
 RESULT 11  
 LOCUS BZ554342  
 DEFINITION BZ554342 1195 bp DNA linear GSS 17-DEC-2002  
 pac61-60\_464.g1 pac61-60 Pseudomonas aeruginosa genomic clone  
 pac61-60\_464, genomic survey sequence.

ACCESSION BZ554342  
 VERSION BZ554342.1 GI:27161567  
 KEYWORDS GSS.  
 SOURCE *Pseudomonas aeruginosa*  
 ORGANISM *Pseudomonas aeruginosa*  
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; *Pseudomonas*.  
 1 (bases 1 to 1195)  
 Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.U., Kaul,R. and Olsen,M.V.  
 Whole-Genome-Sequence Variation among multiple isolates of *Pseudomonas aeruginosa* library  
 J. Bacteriol. (2002) In press  
 Contact: Chris K. Raymond  
 Genome Center  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA  
 Tel: 2062216954  
 Fax: 2066857244  
 Email: craymond@u.washington.edu  
 Class: shotgun.  
 FEATURES  
 source  
 Location/Qualifiers  
 1..1195  
 /organism="Pseudomonas aeruginosa"  
 /mol\_type="genomic DNA"  
 /strain="1-60"  
 /db\_xref="taxon:287"  
 /clone\_lib="pac61-60\_464"  
 /clone\_lib="pac61-60"  
 /note="clinical isolate 1-60 Whole genomic shotgun library."  
 ORIGIN  
 Query Match 10.2%; Score 146.8; DB 9; Length 1195;  
 Best Local Similarity 60.4%; Pred. No. 1.1e-30;  
 Matches 241; Conservative 0; Mismatches 158; Indels 0; Gaps 0;  
 QY 539 GTACCGATTTTTCGCTACCGGCGCTACGCTGTTAAAGCCGAATCTCGGAATTG 598  
 DB 30 GCAAGACTTCGCGCATCTATTCGGCGCCACCTGNAATACCCGAACTGTCGAATTG 89  
 QY 599 AAGCTGTTCGCGTAATATTAAGACCAAGAGATTTGAGCCGCGATGAATGTA 658  
 DB 90 AGACATTCGTTCGCGCTTGGCGCGACGAACCAACTGCTGCGCAAGGCGCGTGA 149  
 QY 659 TTGCGGATTAACGAACTCTCGGCTCTGTTAGTACCGGTTCCGAACAGGATGTGCTGC 718  
 DB 150 TGAACGAATCTGAACTTCGCGGCTGCTGCTGTGACCCGCGCGAGCATGCGATGACCTGCG 209  
 QY 719 TGCAACCGGTAAGAGCGCGCGCTGATATGCAACCAAGGCGAGGAAGTATGACCTTA 778  
 DB 210 TCGGCATTCGCGACGCGCGCTGACATGCGCGCGCGGCGAGGAAGTATGACCTTA 269  
 QY 779 CCGGTGGCGGCGACACGCGTATTTGCGCTTCGCGCGCAACGTCGACGCGGTAATTGCG 838  
 DB 270 CCGGTGGCGGCGATTCGCGTATCTCCACCTTGGCGCGGCGCTTGGCGCGAGAGAC 329  
 QY 839 TGAAGAGCGCTCTCTTTTCCATGCGCGCGCGCTGCGTGTGCGCAAACTGGAA 898  
 DB 330 TGCCTGCGGCGGTGGCGCTGCGCAACCTGCGCGCGCATGCGGCAATCTGGGTA 389  
 QY 899 CTCGACGCTTTCGCGGATGAGACCTGGAATATCTGAC 937  
 DB 390 CCGCGGATCAAGCGCGCGCAACTGCGTGTGCGCGGTGC 428  
 RESULT 12  
 LOCUS BR332786  
 DEFINITION BR332786 505 bp mRNA linear EST 14-JUL-2000  
 u853d03.x1 Perkins LRH Mus musculus cDNA clone IMAGE:3216101 3'  
 similar to SW:RFAB\_ECOLI P76658 ADP-HEPTOSE SYNTHASE ;, mRNA  
 sequence.  
 ACCESSION BR332786

VERSION BE332786.1 GI:206562  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 505)  
 REFERENCE Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Stjepcevic, M., Theising, B., Allen, M., Bowers, Y., Peterson, B., Swaller, T., Giddons, M., Page, D., Harvey, N., Schurk, R., Ritzer, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.  
 The Mashu-NCI Mouse EST Project 1999  
 Unpublished (1999)  
 TITLE Contact: Marra M/Mashu-NCI Mouse EST Project 1999  
 JOURNAL Washington University School of Medicine  
 COMMENT 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty-free through LML, contact the IMRG Consortium (info@imgc.llnl.gov) for further information.  
 MGI:1066249  
 Possible reversed clone: similarity on wrong strand  
 Seq primer: -40RP from Gibco  
 High quality sequence atop: 338.  
 Location/Qualifiers  
 1..505  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="BALB/c"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:3216101"  
 /sex="female"  
 /tissue\_type="primary sorted bone marrow cells"  
 /dev\_stage="adult"  
 /lab\_host="DHI0B (phage-resistant)"  
 /clone\_lib="Perkins LRH"  
 /note="Vector: pZL1, Site\_1: SalI, Site\_2: EagI, cDNA made by oligo-dT priming. Library amplified by stretch PCR. Subtraction method: Bonaldo, et al., Genome Research 6:791. Library constructed by Dr. Archibald Perkins (Yale University)."  
 ORIGIN  
 Query Match 7.0%; Score 100.8; DB 2; Length 505;  
 Best Local Similarity 61.9%; Pred. No. 1.7e-17;  
 Matches 159; Conservative 0; Mismatches 98; Indels 0; Gaps 0;  
 1 ATGAAGTACGCTGCGCAGAGTTTGAACGTGACGAGATGATGATGATGATG 60  
 248 ATGAAGTTCATGATCCCGCATTCATCAAGCCCTGCTGTGTGTGGGATGTCATG 307  
 61 CTGATCGTTACTGTGACGCCCCCAACGATGTAATCTGCGGAAAGCCGCGTCCCGT 120  
 308 CTCGACCGTTATCGGCATGTGTGTGATCTCACGGAATTTCTTGAAGCGCGGTGCTGTG 367  
 121 GTTAAAGTAATACATCATGAAGAAGCTCGGCGCGCGCGCTAACGTGCGATGATATC 180  
 368 GTCAATGTGAGCAATGAGAGATGCCCCGCGTGTGCGGCAAGTTGCCCTTAACATT 427  
 181 GCTTCTCGGAGCTGAATGACGCGCTGATCGGAGTTGACGCGGATTTGACGATGACGCGC 240  
 428 GNCCTCTGGGGGCCCCCGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 487  
 241 GCAGTGAATATCTCT 257  
 488 AGCGTGAACCAAGTCT 504  
 RESULT 13  
 B2648805

LOCUS B2648805 226 bp DNA linear GSS 29-JAN-2003  
 DEFINITION OCCAM49TM ZM 0.7.1.5 KB Zea mays genomic clone ZMMBMA0130J01,  
 genomic survey sequence.  
 ACCESSION B2648805  
 VERSION B2648805.1 GI:28112865  
 KEYWORDS GSS.  
 SOURCE Zea mays  
 ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoidae; Andropogoneae; Zea.  
 1 (bases 1 to 226)  
 REFERENCE White, C.A., Quackenbush, J., Van Aken, S., Uteback, T., Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T., Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.  
 Consortium for Maize Genomics  
 Unpublished (2002)  
 TITLE Other GSSs: OCCAM49TM  
 JOURNAL Contact: Cathy White  
 COMMENT TIGR  
 FEATURES  
 source  
 1..226  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone="ZMMBMA0130J01"  
 /clone\_lib="ZM 0.7.1.5 KB"  
 /note="Vector: pBSCR-; Site\_1: HincII; 0.7-1.5 kb methylation filtered genomic DNA library"  
 ORIGIN  
 Query Match 6.7%; Score 96; DB 9; Length 226;  
 Best Local Similarity 100.0%; Pred. No. 3.5e-16;  
 Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 1123 AACAGCGATGCTCTCAACCAACGCGTGAAGGAGATTCCCGGTAACCACTCGAA 1182  
 131 AACAGCGATGCTCTCAACCAACGCGTGAAGGAGATTCCCGGTAACCACTCGAA 1190  
 1183 CAGCGTATGATTTGTGCTGGCGCGCATCTGGAAGCGCTC 1218  
 191 CAGCGTATGATTTGTGCTGGCGCGCATCTGGAAGCGCTC 226  
 Db  
 191 CAGCGTATGATTTGTGCTGGCGCGCATCTGGAAGCGCTC 226  
 RESULT 14  
 B2648800/c 251 bp DNA linear GSS 29-JAN-2003  
 LOCUS OCCAM49TM ZM 0.7.1.5 KB Zea mays genomic clone ZMMBMA0130J01,  
 genomic survey sequence.  
 ACCESSION B2648800  
 VERSION B2648800.1 GI:28112874  
 KEYWORDS GSS.  
 SOURCE Zea mays  
 ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoidae; Andropogoneae; Zea.  
 1 (bases 1 to 251)  
 REFERENCE White, C.A., Quackenbush, J., Van Aken, S., Uteback, T., Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T., Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.  
 Consortium for Maize Genomics  
 Unpublished (2002)  
 TITLE Other GSSs: OCCAM49TM  
 JOURNAL Contact: Cathy White  
 COMMENT TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843  
Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TP

Class: methylation filtered.

#### FEATURES

Location/Qualifiers

1..251  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZMWBMA0130301"  
/note="Vector: pBCSK-; Site\_1: HincII; 0.7-1.5 kb  
methylation filtered genomic DNA library"

#### ORIGIN

Query Match 6.7%; Score 96; DB 9; Length 251;  
Best Local Similarity 100.0%; Pred. No. 3.5e-16;  
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1123 AACAGCAGTCCCTCCACCAACGGCTGAAGGGGATTCGGCCCGGTAAACCACTCGAA 1182  
DB 96 AACAGCAGTCCCTCCACCAACGGCTGAAGGGGATTCGGCCCGGTAAACCACTCGAA 37  
QY 1183 CAGCGTATGATTGTGCTGGGCGCACTGGAAGCGGTC 1218  
DB 36 CAGCGTATGATTGTGCTGGGCGCACTGGAAGCGGTC 1

#### RESULT 15

CN817146/c

LOCUS CN817146 764 bp mRNA linear EST 18-JUL-2004  
DEFINITION HRO4528 G04 N0825 L1b AA071B1X Avena sativa cv. Og1e-C root Avena

sativa cDNA clone HRO4528\_G04\_N08, mRNA sequence.

CN817146

VERSION CN817146.1 GI:50358506

KEYWORDS EST.

SOURCE Avena sativa (oat)

ORGANISM Avena sativa

REFERENCE Rines, H.W., Anderson, O.D., Crossman, C.C., Lazo, G.R., Miller, S.S.,

AUTHORS Taylor, J.M., and Vance, C.P.

TITLE ESTs from Avena sativa cv. Og1e-C roots, etiolated leaves, and

green leaves

JOURNAL Unpublished (2004)

COMMENT Contact: Rines, H.W.

USDA-ARS and University of Minnesota

1991 Upper Buford Circle, 411 Borlaug Hall, St. Paul, MN 55108, USA

Tel: 612 625 5220

Fax: 612 625 1268

Email: rines001@umn.edu

Sequences have been trimmed to remove low quality sequence with

phred scores less than 20 and most vector sequence.

Seq primer: SK-5prime.

Location/Qualifiers

1..764  
/organism="Avena sativa"  
/mol\_type="mRNA"  
/cultiivar="Og1e-C (a reselection from cv. Og1e)"  
/db\_xref="taxon:498"  
/clone="HRO4528 G04 N08"  
/issue\_type="root"  
/dev\_stage="6-day-old"  
/lab\_host="B. coli SOLR (excised) B. coli TUC121  
(sequencing)"  
/clone\_1b="Lib AA071B1X Avena sativa cv. Og1e-C root"  
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;  
Site\_1: EcoRI; Site\_2: XhoI; Seed were germinated in a

#### ORIGIN

Query Match 6.3%; Score 90.4; DB 7; Length 764;  
Best Local Similarity 58.6%; Pred. No. 1.9e-14;  
Matches 157; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 109 CCGATGCCCGGTGTAAGTAATACCATGAAAGACGTCGGCGCGGCTTAACGTG 168  
DB 764 CCAATGCCCGGTGTAAGTAATACCATGAAAGACGTCGGCGCGGCTTAACGTG 705  
QY 169 GCGATGAATATGCTTCTCTCGGTGTAATGACGCTGCTGTTGACGGGATTTGAC 228  
DB 704 GCTTGAACATTCGCGCGCTGGGCGGACCGCGCTGCTGATCGCGCTCAACCGGCAAGAC 645  
QY 229 GATGACGGCGCGGCTGACTAATCTCGGCGGACGTAAGTCAAGTCAAGTCAAGTCT 288  
DB 644 GAGGCGCGCGGCTGACTAATCTCGGCGGACGTAAGTCAAGTCAAGTCAAGTCT 585  
QY 289 TCTGATCCGACGATCCGACCATTAACCAATTAAGGATCTTCCGCAACCAAGCTG 348  
DB 584 CGATGCGCGGATCCGACCATTAACCAATTAAGGATCTTCCGCAACCAAGCTG 525  
QY 349 ATCCGCTCGGATTTTGAAGAGGTTTCG 376  
DB 524 CTGCGTATGATTTTGAAGAGGTTTCG 497

Search completed: March 18, 2006, 20:52:10  
Job time: 6257 secs

dark incubator at 20 C. At 6 days roots were harvested, total RNA and poly(A) were prepared, a cDNA library was made, and the cDNA clones were in vivo excised to give phagescript phagemids in the Hw Rines and CP Vance USDA-ARS labs (Rines, Vance, Miller, Taylor) at the University of Minnesota, St. Paul, MN. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson USDA-ARS lab (Anderson, Crossman, Lazo) in Albany, CA.

TAG TTSSUB=root  
TAG\_L1b=AA071B1X"

**THIS PAGE LEFT BLANK**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2006, 18:58:16 Search time 449 Seconds  
(without alignments)  
5677.112 Million cell updates/sec

Title: US-10-520-820-28  
Perfect score: 1434  
Sequence: 1 atgaagaatgaacgtgcgcaga.....aacagataaataaagcctaa 1434

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database:

Issued Patents NA: \*  
1: /cgn2\_6/ptodaca/1/ina/1 COMB.seq: \*  
2: /cgn2\_6/ptodaca/1/ina/5 COMB.seq: \*  
3: /cgn2\_6/ptodaca/1/ina/6 COMB.seq: \*  
4: /cgn2\_6/ptodaca/1/ina/6B COMB.seq: \*  
5: /cgn2\_6/ptodaca/1/ina/H COMB.seq: \*  
6: /cgn2\_6/ptodaca/1/ina/PCOMB COMB.seq: \*  
7: /cgn2\_6/ptodaca/1/ina/PP COMB.seq: \*  
8: /cgn2\_6/ptodaca/1/ina/RE COMB.seq: \*  
9: /cgn2\_6/ptodaca/1/ina/backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1434	100.0	1434	US-09-492-709A-165	Sequence 165, App
2	1011.8	70.6	1500	US-09-489-039A-2037	Sequence 2037, Ap
3	870.6	60.7	1323	US-09-489-039A-2107	Sequence 2107, Ap
4	724.2	50.5	1476	US-09-543-681A-2462	Sequence 2462, Ap
5	635.2	44.3	1830121	US-09-557-884-1	Sequence 1, Appli
6	635.2	44.3	1830121	US-09-643-990A-1	Sequence 1, Appli
7	635.2	44.3	1830121	US-10-158-865-1	Sequence 1, Appli
8	494.2	34.5	1581	US-09-252-991A-1613	Sequence 16413, A
9	391	27.3	1251	US-09-252-991A-16104	Sequence 16304, A
10	350.2	24.4	963	US-09-252-991A-16178	Sequence 16178, A
11	238.4	16.6	640681	US-09-790-988-1	Sequence 1, Appli
12	176.2	12.3	501	US-09-252-991A-15975	Sequence 15975, A
13	149.4	10.4	347	US-08-651-155B-185	Sequence 185, App
14	149.4	10.4	347	US-09-194-036B-185	Sequence 185, App
15	144.8	10.1	789	US-09-252-991A-16179	Sequence 16179, A
16	137.6	9.6	9992	US-09-902-540-8926	Sequence 8926, App
17	134	9.3	1442	US-09-902-540-8926	Sequence 8926, App
18	113.6	7.9	58857	US-09-477-962-1	Sequence 1, Appli
19	112.4	7.8	294	US-08-651-155B-185	Sequence 186, App
20	112.4	7.8	294	US-09-194-036B-186	Sequence 186, App
21	100	7.0	312	US-09-252-991A-16071	Sequence 16071, A
22	46.2	3.2	1029	US-09-489-039A-2497	Sequence 2497, Ap
23	45.2	3.2	1840	US-09-902-540-4496	Sequence 4496, Ap
24	45.2	3.2	28320	US-09-902-540-1222	Sequence 1222, Ap

25	41	2.9	1182	3	US-09-270-767-31103	Sequence 31103, A
26	41	2.9	1364	3	US-09-370-767-14850	Sequence 14850, A
27	39.4	2.7	1855	3	US-09-949-016-2757	Sequence 2757, Ap
28	38	2.6	468	3	US-09-252-991A-3080	Sequence 3080, Ap
29	38	2.6	1071	3	US-09-252-991A-2978	Sequence 2978, Ap
30	38	2.6	1488	3	US-09-252-991A-2707	Sequence 2707, Ap
31	37.8	2.6	4403765	3	US-09-489-039A-5586	Sequence 5586, Ap
32	37.4	2.6	4403765	3	US-09-103-840A-2	Sequence 2, Appli
33	37.4	2.6	4411529	3	US-09-103-840A-1	Sequence 1, Appli
34	37	2.6	1818	3	US-09-252-991A-2891	Sequence 2891, Ap
35	37	2.6	7218	2	US-08-232-463-14	Sequence 14, Appli
36	36.8	2.6	1392	3	US-09-902-540-5380	Sequence 5380, Ap
37	36.8	2.6	34552	3	US-09-902-540-1262	Sequence 1262, Ap
38	36.6	2.6	702	3	US-09-303-518B-747	Sequence 747, App
39	36.4	2.5	1551	3	US-09-902-540-4200	Sequence 4200, Ap
40	36.4	2.5	21330	3	US-09-902-540-1209	Sequence 1209, Ap
41	36.4	2.5	4403765	3	US-09-103-840A-2	Sequence 2, Appli
42	36.4	2.5	4411529	3	US-09-103-840A-1	Sequence 1, Appli
43	36	2.5	2326	3	US-09-453-702B-179	Sequence 179, App
44	36	2.5	2326	3	US-10-114-170-179	Sequence 179, App
45	35.8	2.5	459	3	US-09-489-039A-585	Sequence 585, App

## ALIGNMENTS

RESULT 1  
US-09-492-709A-165  
Sequence 165, Application US/09492709A  
Patent No. 6720139  
GENERAL INFORMATION:  
APPLICANT: Zykkind, Judith  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Trawick, John  
APPLICANT: Forsyth, R. Allyn  
APPLICANT: Foelelch, Jamie M.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN  
FILE OF INVENTION: ESCHERICHIA COLI  
FILE REFERENCE: ELITRA, 001A  
CURRENT APPLICATION NUMBER: US/09/492,709A  
CURRENT FILING DATE: 2000-01-27  
NUMBER OF SEQ ID NOS: 485  
SOFTWARE: FASTSEQ for Windows Version 3.0  
SEQ ID NO 165  
LENGTH: 1434  
TYPE: DNA  
ORGANISM: E. Coli  
US-09-492-709A-165  
Query Match 100.0%; Score 1434; DB 3; Length 1434;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGAAGTAACTGTCGACAGTTTGAACGTGACAGAGTATGTTGTTGATGATG 60  
DB 1 ATGAAGTAACTGTCGACAGTTTGAACGTGACAGAGTATGTTGTTGATGATG 60  
QY 61 CTGAGATCGTTACTGTCGACAGGCGCCACCACTGTCGTCGCGGAGCGCGGCGG 120  
DB 61 CTGAGATCGTTACTGTCGACAGGCGCCACCACTGTCGTCGCGGAGCGCGGCGG 120  
QY 121 GTTAAAGTAAATCAATCAAGAAAGCTCGCGGCGCGGCGGCTTAACTGCGATGATATC 180  
DB 121 GTTAAAGTAAATCAATCAAGAAAGCTCGCGGCGCGGCGGCTTAACTGCGATGATATC 180  
QY 181 GCTTCTCTCGGTCATATGACAGCCCTGTCGCTTGAACGCGCATTCATGACAGGCGG 240  
DB 181 GCTTCTCTCGGTCATATGACAGCCCTGTCGCTTGAACGCGCATTCATGACAGGCGG 240  
QY 241 GCGCTGATGAATCTCTGCGCGGACGTCACAGTCAATGCAATTCGTTCTGACGAGG 300  
DB 241 GCGCTGATGAATCTCTGCGCGGACGTCACAGTCAATGCAATTCGTTCTGACGAGG 300

|||||  
Db 241 GCGCTGAGTAAATCTCTGCGCGAAGCTCAAAATGCCACTTCGTTTCTGTACCAAGC 300  
Qy 301 CATCCGACATTAACCAATTAACGGGTACTTTCGCCCAACCAACAGCTGATCCGCTGAT 360  
Db 301 CATCCGACATTAACCAATTAACGGGTACTTTCGCCCAACCAACAGCTGATCCGCTGAT 360  
Qy 361 TTGAAAGAGGTTTCAAGAGTGTGATCCGACGCGCTGCAGAGCGAATTAATCAGGCG 420  
Db 361 TTGAAAGAGGTTTCAAGAGTGTGATCCGACGCGCTGCAGAGCGAATTAATCAGGCG 420  
Qy 421 CTGAGTTGATTTGCGCGCTGCTGCTTTCTGACTACCGCAAGGTGCGCTGCAGAGCTGA 480  
Db 421 CTGAGTTGATTTGCGCGCTGCTGCTTTCTGACTACCGCAAGGTGCGCTGCAGAGCTGA 480  
Qy 481 CAGAGATGATCCAACTGCGCGGTAAAGCGGGTTCCTCGGCTGATTAATCCAAAAGT 540  
Db 481 CAGAGATGATCCAACTGCGCGGTAAAGCGGGTTCCTCGGCTGATTAATCCAAAAGT 540  
Qy 541 ACCGATTTTGAAGCGCTACCGCGCGCTACCGCTGTTAAACCGCAATCTCTCGAATTTGAA 600  
Db 541 ACCGATTTTGAAGCGCTACCGCGCGCTACCGCTGTTAAACCGCAATCTCTCGAATTTGAA 600  
Qy 601 GCTGTGTGCGTAAATGTAGACCGAAGAGAGATTGTAAGCGCGCGATGAATCTGATT 660  
Db 601 GCTGTGTGCGTAAATGTAGACCGAAGAGAGATTGTAAGCGCGCGATGAATCTGATT 660  
Qy 661 GCGCAATTAACGAATCTGCGGCTCTGTGATGACCGGTTCCGAACAGGGTATGTCGCTG 720  
Db 661 GCGCAATTAACGAATCTGCGGCTCTGTGATGACCGGTTCCGAACAGGGTATGTCGCTG 720  
Qy 721 CAACCGGGTAAAGCGCGCTCATATGCAACCAACGCGCAGAGATGTATGACGTTACC 780  
Db 721 CAACCGGGTAAAGCGCGCTCATATGCAACCAACGCGCAGAGATGTATGACGTTACC 780  
Qy 781 GGTGCGGGCGACACGCTGATTTGCGTCTGCGCGCAACGCTGCAAGCGGGTAAATTCCTG 840  
Db 781 GGTGCGGGCGACACGCTGATTTGCGTCTGCGCGCAACGCTGCAAGCGGGTAAATTCCTG 840  
Qy 841 GAAAGAGCTGCTCTTTGCCAATGCGCGCGCTGCGGTGCGGTGCGCAACTGCGAACC 900  
Db 841 GAAAGAGCTGCTCTTTGCCAATGCGCGCGCTGCGGTGCGGTGCGCAACTGCGAACC 900  
Qy 901 TCACAGGTTTGCAGATCGAGCTGGAATAATGCTGTACGTGACGTGACATACAGGCTTT 960  
Db 901 TCACAGGTTTGCAGATCGAGCTGGAATAATGCTGTACGTGACGTGACATACAGGCTTT 960  
Qy 961 GCGCTGATGACCGAAGAGAACTGAAGCTGCGCGGTAGCGCGCAAGCTGTGAA 1020  
Db 961 GCGCTGATGACCGAAGAGAACTGAAGCTGCGCGGTAGCGCGCAAGCTGTGAA 1020  
Qy 1021 AAAGTGTGATGACCAACGAGTGTCTTTGACATCCGACGCGCGGCAAGCTGTCTTATCTG 1080  
Db 1021 AAAGTGTGATGACCAACGAGTGTCTTTGACATCCGACGCGCGGCAAGCTGTCTTATCTG 1080  
Qy 1081 GCAAAATGCGCGCAAGCTGAGCTGCTGATTTGATTTGCGGTCAACAGCGATGCTTCAACC 1140  
Db 1081 GCAAAATGCGCGCAAGCTGAGCTGCTGATTTGATTTGCGGTCAACAGCGATGCTTCAACC 1140  
Qy 1141 AAAGCGCTGAAAGGGGATTTCCGCCCGGTAAACCACTCGAACAGCGTATGATGTGCTG 1200  
Db 1141 AAAGCGCTGAAAGGGGATTTCCGCCCGGTAAACCACTCGAACAGCGTATGATGTGCTG 1200  
Qy 1201 GGGCACTGGAAGGGGTGACATGGGTAGTGTGTTTGAAGAGGACAGCGCGCAAGCGCTG 1260  
Db 1201 GGGCACTGGAAGGGGTGACATGGGTAGTGTGTTTGAAGAGGACAGCGCGCAAGCGCTG 1260  
Qy 1261 ATCCCGGGGATCTTGCCAGATCTGCTGTGTAAGGCGCGCATTAACCAAGAGAGATT 1320  
Db 1261 ATCCCGGGGATCTTGCCAGATCTGCTGTGTAAGGCGCGCATTAACCAAGAGAGATT 1320  
Qy 1321 GCGCGAGTAAAGAGTCTGCGGCAACGCTGCGCAAGCTTGTGCTCACTTTGAAGAC 1380  
Db 1321 GCGCGAGTAAAGAGTCTGCGGCAACGCTGCGCAAGCTTGTGCTCACTTTGAAGAC 1380

Db 1321 GCGCGAGTAAAGAGTCTGCGGCAACGCTGCGCAAGCTTGTGCTCACTTTGAAGAC 1380  
Qy 1381 GGTGTGCGACGACCAACATCATCAAGAGATCCACAGGATAAAGGCTAA 1434  
Db 1381 GGTGTGCGACGACCAACATCATCAAGAGATCCACAGGATAAAGGCTAA 1434  
RESULT 2  
US-09-489-039A-2037  
; Sequence 2037, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489, 039A  
; PRIOR FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 2037  
; LENGTH: 1500  
; TYPE: DNA  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-2037  
Query Match 70.6%; Score 1011.8; DB 3; Length 1500;  
Best Local Similarity 81.9%; Pred. No. 5e-285;  
Matches 1166; Conservative 0; Mismatches 257; Indels 0; Gaps 0;  
Qy 1 ATGAAAGTAAAGCTGCGCAGAGTTTGAACGTGACAGAGTGTGTTGATGATGATG 60  
Db 67 ATGAAAGTAAAGCTGCGCAGAGTTTGAACGTGACAGAGTGTGTTGATGATGATG 126  
Qy 61 CTGAGTGTACTGTGTAACGCGCCCAACAGTCTGATCTCGCGGAAAGCGCGTCCGCTG 120  
Db 127 CTGAGTGTACTGTGTAACGCGCCCAACAGTCTGATCTCGCGGAAAGCGCGTCCGCTG 186  
Qy 121 GTTAAAGTAAATCAATCGAAGAAAGTCCGCGCGCGCGCTAAACGTGCGATGATATC 180  
Db 187 GTTAAAGTAAATCAATCGAAGAAAGTCTGCGCGCGCGCGCTAAACGTGCGATGATATC 246  
Qy 181 GCTTCTCTCGGTCATATGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240  
Db 247 GCTTCTCTCGGTCATATGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 306  
Qy 241 GCGCTGATTAATCTGCGCGCAAGCTGCAAGCTCAATGCAATGCAATCTTCTGTAACG 300  
Db 307 GCGCTGATTAATCTGCGCGCAAGCTGCAAGCTCAATGCAATGCAATCTTCTGTAACG 366  
Qy 301 CATCCGACATTAACCAATTAACGGGTACTTTCGCCCAACCAACAGCTGATCCGCTGATC 360  
Db 367 CATCCGACATTAACCAATTAACGGGTACTTTCGCCCAACCAACAGCTGATCCGCTGATC 426  
Qy 361 TTGAAAGAGGTTTCAAGAGTGTGATCCGACGCGCTGCAGAGCGGATTAATCAGGCG 420  
Db 427 TTGAAAGAGGTTTCAAGAGTGTGATCCGACGCGCTGCAGAGCGGATTAATCAGGCG 486  
Qy 421 CTGAGTTGATTTGCGCGCTGCTGCTTTCTGACTACCGCAAGGTGCGCTGCAGAGCTGA 480  
Db 487 CTGAGTTGATTTGCGCGCTGCTGCTTTCTGACTACCGCAAGGTGCGCTGCAGAGCTGA 546  
Qy 481 CAGAGATGATCCAACTGCGCGGTAAAGCGGGTTCCTCGGCTGATTAATCCAAAAGT 540  
Db 547 CAGAGATGATCCAACTGCGCGGTAAAGCGGGTTCCTCGGCTGATTAATCCAAAAGT 606  
Qy 541 ACCGATTTTGAAGCGCTACCGCGCGCTACCGCTGTTAAACCGCAATCTCTCGAATTTGAA 600  
Db 607 ACCGATTTTGAAGCGCTACCGCGCGCTACCGCTGTTAAACCGCAATCTCTCGAATTTGAA 666  
Qy 601 GCTGTGTGCGTAAATGTAGACCGAAGAGAGATTGTAAGCGCGCATGAATCTGATT 660  
Db 601 GCTGTGTGCGTAAATGTAGACCGAAGAGAGATTGTAAGCGCGCATGAATCTGATT 660



```

Db      667 GCGGTGTGGGCAAGTGCAGATGAAGCGCAGATGCTTGAGCGCGCATGAAGCTCATC 726
Qy      661 GCGGATTAAGAACTCTGCGCTCTGTTAGTGAACCGCTTCCGAACAGGGTATGTGCTGCTG 720
Db      727 GCGGAATTCACACTGTGCGCGCTGCTGTGTACCGGCTCTAGACAGGGGATGAAGCTGCTG 786
Qy      721 CAACCGGGTAAAGCGCGCTGCAATATGCAACCCAGCGAGAGAGTATGACGTAAACC 780
Db      787 CAGCGGAGACGTCCGCGCTGCAATATGCAACCCAGCGAGAGAGTATGACGTAAACC 846
Qy      781 GGTGCGGGGCAACCGTGAATGCGCTCTGCGCGCAACCGTGGCAGCGGGTAAATTCGCTG 840
Db      847 GGGCGCGGCAACCGTGAATGCGCTCTGCGCGCAACCGTGGCAGCGGGTAAATTCGCTG 906
Qy      841 GAAGAAAGCTCTCTCTTCTGCAATGCGCGCGCTGCGCTGCGTGGTAACTGGGAAAC 900
Db      907 GAAGAAAGCTCTCTCTTCTGCAATGCGCGCGCTGCGCTGCGTGGTAACTGGGTAACA 966
Qy      901 TCACGCGTTCCGCGATCGAGCTGAAGAAATGCTGTATGCTGAGACGTGCAATACAGCTTT 960
Db      967 TCACGCGTTCCGCGATCGAGCTGAAGAAATGCTGTATGCTGAGACGTGCAATACAGCTTT 1026
Qy      961 GCGGTGATGACCGAAGAGAACTGAAGCTGCGCTAGCGCGCAGCGCGTAAACGTGTGA 1020
Db      1027 GCGGTGATGACCGAAGAGAACTGAAGCTGCGCTAGCGCGCAGCGCGTAAACGTGTGA 1086
Qy      1021 AAAGTGTATGACCAACGCTGTCTTGAATCTCTGACCGCGCGACGCTCTTATCTG 1080
Db      1087 AAAGTGTATGACCAACGCTGTCTTGAATCTCTGACCGCGCTCTTATCTG 1146
Qy      1081 GCAATGCGCGCAGCTGCGGTGACCGCTGATTTGTGCGCGTCAACAGCGATGCTCCACC 1140
Db      1147 GCGAATGCGCGCAGCTGCGGTGACCGCTGATTTGTGCGCGTCAACAGCGATGCTCCACC 1206
Qy      1141 AAACGCGTGAAGAGGAAATCCCGCCCGCTAAACCACTCGAACAGCTATGATGCTG 1200
Db      1207 AAACGCGTGAAGAGGAAATCCCGCCCGCTAAACCACTCGAACAGCTATGATGCTG 1266
Qy      1201 GCGGCACTGGAAGCGGTGACCTGCTGATGCTGTTGAAGAGACAGCGCGACGCTTG 1260
Db      1267 GCGGCGCTGGAAGCGGTGACCTGCTGATGCTGTTGAAGAGATCTCCGACGCGCTG 1326
Qy      1261 ATCGCGCGGATCTTGGCAGATCTGCTGTTGAAGAGCGCGCACTATTAACAGAGATG 1320
Db      1327 ATCGCTGCGATCTCTGCGGATCTGCTGTTGAAGAGCGCGCACTATTAACAGAGATG 1386
Qy      1321 GCGCGAGTAAAGAGTCTGCGCGCAACCGTGGCGAAGTGTGTCTCAACTTTGAAGAC 1380
Db      1387 GCGCGAGTAAAGAGTCTGCGCGCAACCGTGGCGAAGTGTGTCTCAACTTTGAAGAT 1446
Qy      1381 GGTGTCTGACGACCAATCATCAGAGATCCACAGATTA 1423
Db      1447 GGTGTCTGACCAATCATCAGAGATTAAGAAATTAAGAAATTA 1489

```

```

RESULT 3
US-09-489-039A-2107/c
/ Sequence 2107, Application US/09489039A
/ Patent No. 6610836
/ GENERAL INFORMATION:
/ APPLICANT: Gary Breton et. al
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
/ FILE REFERENCE: 2709.2004001
/ CURRENT APPLICATION NUMBER: US/09/489,039A
/ PRIOR APPLICATION NUMBER: 2000-01-27
/ PRIOR FILING DATE: 1999-01-29
/ NUMBER OF SEQ ID NOS: 14342
/ SEQ ID NO 2107
/ LENGTH: 1323
/ TYPE: DNA
/ ORGANISM: Klebsiella pneumoniae

```

```

US-09-489-039A-2107
Query Match      60.7%; Score 870.6; DB 3; Length 1323;
Best Local Similarity 81.8%; Pred. No. 8.1e-244;
Matches 1005; Conservative 0; Mismatches 224; Indels 0; Gaps 0;

Qy      1 ATGAAGTAAAGCTGCGCAAGGTTGAACGTGACGAGAGTATGCTGTGATGATG 60
Db      1230 ATGAAGTAAAGCTGCGCAAGGTTGAACGTGACGAGAGTATGCTGTGATGATG 1171
Qy      61 CTGAATCGTTACTGTAGCGGCCCAACAGTGTATCTGCGCGAAGCGCGGCTG 120
Db      1170 CTGAACCGCTACTGTAGCGGCCCAACAGTGTATCTGCGCGAAGCGCGGCTG 1111
Qy      121 GTTAAAGTAAATCAATCGAAGAAAGTCCGCGCGCGCGCTTAAAGTGGCGATGATTC 180
Db      1110 GTTAAAGTAAATCAATCGAAGAAAGTCCGCGCGCGCGCTTAAAGTGGCGATGATTC 1051
Qy      181 GCTTCTCTGCGTCTAATGACAGCTGTGTCGAGTTGACGCGGCAATTGACGAGCGCG 240
Db      1050 GCTTCTCTGCGTCTAATGACAGCTGTGTCGAGTTGACGCGGCAATTGACGAGCGCG 991
Qy      241 GCGCTGATGAATCTCTGCGCGCAAGTCAACGTCMAATGCGACTTCTGTAACGACG 300
Db      990 GCGCTGATGAATCTCTGCGCGCAAGTCAACGTCMAATGCGACTTCTGTAACGACG 931
Qy      301 CATCGACCAATTAACAAATTAAGGGTACTTCCCGCAACCAACGCTGATCCGCTGAT 360
Db      930 CATCGACCAATTAACAAATTAAGGGTACTTCCCGCAACCAACGCTGATCCGCTGAT 871
Qy      361 TTGGAAGAGGTTTGAAGAGTGTGATTCGACAGCGCTGCAACAGCGGATTAATCAGCG 420
Db      870 TTGGAAGAGGTTTGAAGAGTGTGATTCGACAGCGCTGCAACAGCGGATTAATCAGCG 811
Qy      421 CTGAGTTCGATTTGCGCGCTGCTGCTTCTGATCAACGCAAGGTCGCTGCAACGCTA 480
Db      810 CTGAGTTCGATTTGCGCGCTGCTGCTTCTGATCAACGCAAGGTCGCTGCAACGCTA 751
Qy      481 CAGCAATGATCCAACTGCGCGCTAAAGGGGTGTTCCGCTGCTGATTTGATCCAAAGT 540
Db      750 CAGCAATGATCCAACTGCGCGCTAAAGGGGTGTTCCGCTGCTGATTTGATCCAAAGT 691
Qy      541 ACCGATTTGAGCGCTACAGCGCGCGCTGCTGCTGTTAAACGCGCAATCTCTCGAATTGAA 600
Db      690 ACCGATTTGAGCGCTACAGCGCGCGCTGCTGCTGTTAAACGCGCAATCTCTCGAATTGAA 631
Qy      601 GCTGTGTGCGTAAATGTAAGACCGAAGAGATGTTGAGCGCGCATGAAGTGAAT 660
Db      630 GCTGTGTGCGTAAATGTAAGACCGAAGAGATGTTGAGCGCGCATGAAGTGAAT 571
Qy      661 GCGGATTAAGAACTCTGCGCTCTGTTAGTGAACCGTTCCGAACAGGTTATGTGCTGCTG 720
Db      570 GCGGATTAAGAACTCTGCGCTCTGTTAGTGAACCGTTCCGAACAGGTTATGTGCTGCTG 511
Qy      721 CAACCGGGTAAAGCGCGCTGCAATATGCAACCCAGCGAGAGAGTATGACGTAAACC 780
Db      510 CAACCGGGTAAAGCGCGCTGCAATATGCAACCCAGCGAGAGAGTATGACGTAAACC 451
Qy      781 GGTGCGGGGCAACCGTGAATGCGCTCTGCGCGCAACCGTGGCAGCGGGTAAATTCGCTG 840
Db      450 GGTGCGGGGCAACCGTGAATGCGCTCTGCGCGCAACCGTGGCAGCGGGTAAATTCGCTG 391
Qy      841 GAAGAAAGCTCTCTCTTCTGCAATGCGCGCGCTGCGCTGCGTGGTAACTGGGAAAC 900
Db      390 GAAGAAAGCTCTCTCTTCTGCAATGCGCGCGCTGCGCTGCGTGGTAACTGGGTAACA 331
Qy      901 TCACGCGTTCCGCGATCGAGCTGAAGAAATGCTGATGCTGACGTCAGATACAGGCTTT 960
Db      330 TCACGCGTTCCGCGATCGAGCTGAAGAAATGCTGATGCTGACGTCAGATACAGGCTTT 271
Qy      961 GCGGTGATGACCGAAGAGAACTGAAGCTGCGCTGAGCGCGCAAGCGGTGAACGTGTGA 1020
Db      270 GCGGTGATGACCGAAGAGAACTGAAGCTGCGCTGAGCGCGCAAGCGCGGTGAACGTGTGA 211

```





Owen White  
Hamilton O. Smith  
J. Craig Venter  
TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments thereof, and Uses Thereof  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville,  
STATE: MD  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3 1/2 inch diskette  
COMPUTER: Dell Pentium  
OPERATING SYSTEM: MS DOS v6.22  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/643,990A  
FILING DATE: 23-Aug-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/487,429  
FILING DATE: 1995-06-07  
APPLICATION NUMBER: 08/426,787  
FILING DATE: 1995-04-21  
ATTORNEY/AGENT INFORMATION:  
NAME: Kenley K. Hoover  
REGISTRATION NUMBER: 40,302  
REFERENCE/DOCKET NUMBER: PB186P1C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 301-610-5790  
TELEFAX: 310-309-8439  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1830121 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-643-990A-1  
Query Match 44.3%; Score 635.2; DB 3; Length 1830121;  
Best Local Similarity 65.9%; Pred. No. 1e-173;  
Matches 922; Conservative 0; Mismatches 478; Indels 0; Gaps 0;  
QY 17 CAGAGTTGAACGTGCAGAGTGAATGTTGGTGTGATGTGATCTGGATCTTACTGGT 76  
DB 1596554 CAGATTTAACGACAAAGAACTACTCTGATTAGCGACGTGATCTGATCTGTTATTGGT 1596495  
QY 77 ACGGCCCCACGACGTCTATCTCCCGGAAAGCGCGGTCCCGTGTAAAGTGAATACA 136  
DB 1596494 TCGGCGCAACCAACCGTATTTACAGAAAGCAAGTACCAAGTGGTCTGTGTACAGAAA 1596435  
QY 137 TCGAAGAACTCCGCGCGCGCGCTAACTGCGAGTGAATATGCTTCTCGGTGCTA 196  
DB 1596434 ATGAAGAAACGCGCGGTGTGTCAAGCAATGTGCGAGTGAATATGCTTCACTCAATGTAC 1596375  
QY 197 ATGACGCGCTGGTGGGTGTAACGCGCATTTGACATGACGCGCGCGCTGAATATCTC 256  
DB 1596374 CCGTTCAGTTAATGGGATTTGATTGACAAAGATGAATGATGCTTCTGCACTTCCCTCTTAT 1596315  
QY 257 TGGCGCAGTCAAGTCAAAATGCGACTTCCTTTCTGTATCCGACGCAATCCGACCAATTACA 316  
DB 1596314 TAGAAAAACAAAATGATTGTATTTTGTGCAATTAGAAAACCATCAACCAATTACTA 1596255  
QY 317 AATTACGGGTACTTTCGCGAACCAACAGTATCCGCTGCTGATTTTGAAGAAGTTTGG 376  
DB 1596254 AATTACGATTTTATCTGTATCAACAGCTGCTCCGCTTGAATTTTGAAGAATTTTCA 1596195  
\* QY 377 AAGGTGTTGATCCGACGCGCTGCAAGAGCGATTTAATCAAGCGCTGAATTTGATGGCG 436

DB 1596194 AATATGATGATGCAAGATTTTATTAGCGAAGTTAGAAAAGTGGCGTGAATAAATTACGGTG 1596135  
QY 437 CGGTGATGCTTTTGTGCTACGCCAAGGTGGCGTGGCAAGCGATGATCAAC 496  
DB 1596134 CTTTGATTTCTTTGTGATTACGGCAAGGCAAGCTTTAAAGATTTCAGAAATATTTCAAA 1596075  
QY 497 TGGCGGTAAAGCGGGGTGTTCCGCTGTATGATCAAAAAGTACCGATTTTGAAGCGCT 556  
DB 1596074 TTGCAGCGAAAGGAATGTGCTGTGTATGATCAAAAGGAATCGATTTTGAAGCTT 1596015  
QY 557 ACCGCGCGCTACCTGTTTAAAGCCGCAATCTCTCGAATTTGAAGCTTTGTGCTTAAT 616  
DB 1596014 ATCGTGGGCTACATTATTGACACCCCAATATGCTCAATTTGAAGCTTTGTGGTAAAT 1595955  
QY 617 GTAAAGCCGAAGAAAGATTTGTTAGCGGGCAATGAACCTGATTTGCGATTAAGAACTCT 676  
DB 1595954 GCAATACGGAAGAAAGATTTATTAAGAAAGTTTAAATTTATTTCTGATATTGAATTA 1595895  
QY 677 CGGCTGTGTTAGTGAACCGCTTCCGAAACAGGTATGTCTGCTGCAACCGGGTAAAGCGC 736  
DB 1595894 CGCACCTTTTGTGAGCGCGTTCGAAAGGACATGACATTATTACGCCCAATCAAGAGC 1595835  
QY 737 CGCTGCATATGCCAACCCAAAGCGCAGAAAGTATGACGTTACCGGTGGCGGCGACACGG 796  
DB 1595834 CTTATCATTTTGCACACGTTGCAAAAAGATGTTTATGTATGTAGACGGAGCGGTGACACTG 1595775  
QY 797 TGAATGGCGCTCTGGCGGCAACGCTGCAAGCGGGTAAATTCGTGAAGAAAGCTGCTTCT 856  
DB 1595774 TCAATTAGCGTATTATGCAACCGCATTTAGCAGATGAGAGTCTTTTCGAAGAACTTTGTTACC 1595715  
QY 857 TTGCAATGCGGCGGCTGGCGGTGTGTGTGCGCAAACTGGGAACCTCCAGGTTTTCGCGA 916  
DB 1595714 TAGCAATGTGTGCGCAGGAATGTGTGTGGGAAATTTGGGGAATTCAACGGTTTTCACCG 1595655  
QY 917 TCGAGCTGGAATAATGCTGTATGCTGTGACGTCGACATACAGGCTTTTGGCGGTATGACGGAAG 976  
DB 1595654 TGGAACTTGAAATGTGGAATTCAGTCTGCTGTGAACTGTGATTTGCAATTATAGTGAAG 1595595  
QY 977 AGGAATGAAGCTGGCGCGTGAACGCGGCAAGCGGTAAAGTGTGTAAGAAAGTGTATGACCA 1036  
DB 1595594 CAGAAATTAAGAAATGCTGTGCGCAACAGCTTAAGCGCGGTGAATAATTTGTATGACTTA 1595535  
QY 1037 ACGGTGCTTTGACATCTGACAGCGCGGCAAGCTCTTATCTGGCAAAATGCGCGCAAGC 1096  
DB 1595534 ATGGCGTTTGTGATTTTGTGACATCAAGGGCATATTTCTTATTTAGAAAATGACAGCAAT 1595475  
QY 1097 TGGGTGACGCTTGATTTGTGCGGTCAACAGCGATGCTTCAACAAACGGCTGAAAGGGG 1156  
DB 1595474 TGGGCGATTCCTTAATTTGTGCGGTAAACAGCGACATTCCTGTAAACGCTTAAAGGTG 1595415  
QY 1157 ATTCCGCGCGGTAAACCCCACTCGAAACAGCGTATGATTTGTGCGCGCATGGAAGCGG 1216  
DB 1595414 AAAGTGCCCAATTAATTAATCTTGAAGAAACCGTATGCGGATTTGGCTGTGTTGGATCCG 1595355  
QY 1217 TCGATGCGGTATGTGCTTTGAAAGAGACACGCGCAAGCGCTTATCGCGGGAATCTTGC 1276  
DB 1595354 TAGACTGTTGTGTGCTTTCATGAAAGATACCAACACGTTTAATTCGGGAATTTCTAC 1595295  
QY 1277 CAGATCTGCTGTGAAGAGGCGGACATTAATACAGAAAGATTTGCCGGAATGAAGAG 1336  
DB 1595294 CAGATCTTTTATGCAAAAGGCGGATTTACAAACCGAAGATTTGACGAGGATGAAGAG 1595235  
QY 1337 TCTGGGCAACGCTGGCGAAGGTGTGTGCTCAACTTTGAAAGCGGTGCTGACAGCA 1396  
DB 1595234 TTTGGGCAACGCTGGCGATGTTAAAGTGTCTTAACCTTTGAATAATGTTGTTCAACAA 1595175  
QY 1397 ACATCATCAAGAAATCAAA 1416  
DB 1595174 ATGTGATTAAGAAAATTTAA 1595155

RESULT 7

US-10-158-865-1/c  
Sequence 1, Application US/10158865  
Patent No. 684651  
GENERAL INFORMATION:  
APPLICANT: Pleischmann et al.  
TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome, Frag  
Patent No. 684651  
TITLE OF INVENTION: Thereof, and Uses Thereof  
FILE REFERENCE: P186P2CID1  
CURRENT APPLICATION NUMBER: US/10/158,865  
PRIOR FILING DATE: 2002-06-03  
PRIOR APPLICATION NUMBER: US 09/557,884  
PRIOR FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: US 08/476,102  
PRIOR FILING DATE: 1995-06-07  
PRIOR APPLICATION NUMBER: US 08/426,787  
PRIOR FILING DATE: 1995-04-21  
NUMBER OF SEQ ID NOS: 1  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 1  
LENGTH: 1830121  
TYPE: DNA  
ORGANISM: Haemophilus influenzae  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (4747)..(4747)  
OTHER INFORMATION: n equals a,t,c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (9921)..(9921)  
OTHER INFORMATION: n equals a,t,c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (10150)..(10150)  
OTHER INFORMATION: n equals a,t,c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (29298)..(29298)  
OTHER INFORMATION: n equals a,t,c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (36543)..(36543)  
OTHER INFORMATION: n equals a,t,c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (36551)..(36551)  
OTHER INFORMATION: n equals a,t,c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (36636)..(36636)  
OTHER INFORMATION: n equals a,t,c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (40808)..(40810)  
OTHER INFORMATION: n equals a,t,c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (44416)..(44416)  
OTHER INFORMATION: n equals a,t,c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (44905)..(44905)  
OTHER INFORMATION: n equals a,t,c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (44975)..(44975)  
OTHER INFORMATION: n equals a,t,c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (45593)..(45593)  
OTHER INFORMATION: n equals a,t,c, or g  
FEATURE:  
NAME/KEY: misc\_feature

LOCATION: (45732)..(45732)  
OTHER INFORMATION: n equals a,t,c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (47036)..(47036)  
OTHER INFORMATION: n equals a,t,c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (51334)..(51334)  
OTHER INFORMATION: n equals a,t,c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (51602)..(51602)  
OTHER INFORMATION: n equals a,t,c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (51786)..(51786)  
OTHER INFORMATION: n equals a,t,c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (51805)..(51805)  
OTHER INFORMATION: n equals a,t,c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (55369)..(55369)  
OTHER INFORMATION: n equals a,t,c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (65309)..(65309)  
OTHER INFORMATION: n equals a,t,c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (65313)..(65313)  
OTHER INFORMATION: n equals a,t,c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (80024)..(80024)  
OTHER INFORMATION: n equals a,t,c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (100091)..(100091)  
OTHER INFORMATION: n equals a,t,c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (102696)..(102696)  
OTHER INFORMATION: n equals a,t,c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (105121)..(105121)  
OTHER INFORMATION: n equals a,t,c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (107248)..(107248)  
OTHER INFORMATION: n equals a,t,c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (117136)..(117136)  
OTHER INFORMATION: n equals a,t,c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (119750)..(119750)  
OTHER INFORMATION: n equals a,t,c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (119924)..(119924)  
OTHER INFORMATION: n equals a,t,c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (120038)..(120038)  
OTHER INFORMATION: n equals a,t,c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (121344)..(121344)

```
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (13910)..(13910)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152500)..(152500)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
NAME/KEY: misc_feature
Query Match 44.3%; Score 635.2; DB 3; Length 1830121;
Best Local Similarity 65.9%; Pred. No. 1e-173;
Matches 922; Conservative 0; Mismatches 478; Indels 0; Gaps 0;
QY 17 CAGAGTTTGAACGTGACGAGTGTGTTGTGTGATGATGCTGATGTTACTGTT 76
DB 1596554 CAGATTTTAACGACAAAGTAATCTGATTAGGCGAGTATGATCTGATTGTT 1596495
QY 77 ACGGCCCCACGACGTCTCTCGCGGAAAGCCCGGTGCCGTGTTAAATGAAATCA 136
DB 1596494 TCGGCGCAACCAACCGTATTTTCAACGAAAGCAACGATGCTGTTGTAACAAGAA 1596435
QY 137 TCGAAGAACTGTCGGGCGGCGGCTTAACGTGCGATGATATGCTTCTCGGTGCTA 196
DB 1596434 ATGAAGAAACGGCGGGGTGTGTCAGCAAAATGTGCGGATGAATATGCTTCACTCAATGTA 1596375
QY 197 ATGCAACGCTGTGCGGTTGACGGGCAATTGACGATGACGCGCGCTGTAATCTC 256
```

```
DB 1596374 CCGTTAGTTAATGGATTCATGTCAGCAAGATGAACTGGTCTGCACTTCCCTCTAT 1596315
QY 257 TGGCCGACGTCAACGTCAATGCGACTTGTTCTGTACCGACGATCCGACATTACCA 316
DB 1596314 TAGAAAAAATAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1596255
QY 317 AATTACGGGTACTTTCCTCCGCAACCAAGCTGATCCGTGATTTTGAAGAAGTTTCG 376
DB 1596254 AATTACGTTATTTATCTCGATCAACAGCTGCTCCGCTGATTTTGAAGAATTTCA 1596195
QY 377 AAGGTGATCCGACCGGCTGACAGAGGATTAATCAGGCGCTGATTCATTTGCGC 436
DB 1596194 ATATGATGATTCGCAAGATTTTATAGCGAAGTTAGAAAGTGGCGGTAATAATACG 1596135
QY 437 CGCTGATCTTTGCTGCTACCGCAAGGTGCGTGCAGACGCTGATGATCCAAAC 496
DB 1596134 CTTGATCTTTCTGATTAACGGCAAGGCAACGCTTAAGATGTTCAAGAAATATTCAAA 1596075
QY 497 TGGCGGTAAAGCGGGGTGTTCCGCTGATGATGATCAAAAGGTACCGATTTTGAAGCT 556
DB 1596074 TTGCAAGCAAGGCAATGTGCTGTGTTGATGATCAAAAGGCACTGATTTGAAGCT 1596015
QY 557 ACCGCGGCGCTACGCTGTTAACCCGCAATCTCTCGAATTTGAAGCTGTTGCTTAAT 616
DB 1596014 ATGCTGGGCTACATTATGACACCAATATGCTGAATTTGAAGCTGTGTGCTTAAT 1595955
QY 617 GTAAAGCCGAAGAGATTTGTTAGCGCGGCACTGAACCTGATTTGCGATTAACCTCT 676
DB 1595954 GCAATACGAAAGAGATTTATGAAAGGTTTAAATTTTGTGATTTGAATTTAA 1595895
QY 677 CGGCTCTGTTAGTGAACCGTTCGAAACAGGTATGTCTGCTGCAACCGGGTAAAGCGC 726
DB 1595894 CGCACCTTTTGTGAGCGGTTTGAAAGGCAATGACATTATTAACGCCCAATCAAGAGC 1595835
QY 737 CGCTGATATGCAACCCAGACGCAAGGTATGACGTTACCGGTGGGGGCAACCG 796
DB 1595834 CTATATCTTGGCAACTGTTGCAAAAGAGTGTATGTGAGCGGAGCGGTGACACTG 1595775
QY 797 TGAATGGCGCTCCGCGGCAACGCTGCGAGCGGTTAATCGCTGGAAGAACCTGCTCT 856
DB 1595774 TCAATTAGCTATTATGCAACCGCAATTAGCAATGAGAGCTTTTCAAGAAATCTTTTACC 1595715
QY 857 TTGCAATGCGCGGCGCTGCGGTGTGTGCTGCGCAACTGGGAACCTCAGCGTTTCGCG 916
DB 1595714 TAGCAATGTTGCGGAGGAATGTGTGTGGGAATTTGGGCACTTCAACGTTTCGACCG 1595655
QY 917 TCGAGCTGGAATAATGCTGATGACGTGACGTGACATACAGGCTTTGGCGTATGACGAAG 976
DB 1595654 TGGAACTTGAAATGCGATTCATGCTGCTGAAACTGGAATTTGCAATTATGAGTGAAG 1595595
QY 977 AGGAACTGAAGCTGGCGGTGAGCGGCAACGCGGTAAAGCTGTTGAAGAAAGTGTATGACCA 1036
DB 1595594 CAGATTTAAAGATGCTGTGCGCAACAGCTTAAGCGCGGTGAAGAAATTTGTATGACTA 1595535
QY 1037 ACGGTGCTTTGACATCTGACAGCGCGGCAACGCTCTTATCTGCAAAATACCGCAAGC 1096
DB 1595534 ATGGCTGTTTATTTATTTTGGCAATCCAGGGGATATTTCTTATTTAAGAAATGACGCAAT 1595475
QY 1097 TGGGTGACGCTGATTTGTTGCGGTCAACAGCATGCTTCAACAAACGCGTGAAGGGG 1156
DB 1595474 TGGGCAATGCTTAATTTGTTGCGGTAAACAGCACCAATCTGTAAACGCTTAAAGGTG 1595415
QY 1157 ATTCCGCGCGGTAAACCACTGGAACAGGTATGATTTGCTGGGCGCACTGGAAGCGG 1216
DB 1595414 AAGTGCGCCCAATTAATTAATCTTGAAGAAACGTAATGGCGGTATTTGGCTGTTTGGCATCCG 1595355
QY 1217 TCGACTGAGTATGTGTTTGAAGAGACACGCGCAACGCTTGAATCGCGGGAATCTTGC 1276
DB 1595354 TAGACTGTTTGTGCTTTCACTGAAGATACCAACAGTTTAATTCGGGGAATTTCTAC 1595295
QY 1277 CAGATCTGCTGTGAAGAGCGGCACTATTAACCAAGAGATTTGCCGGAGTAAAGAG 1336
```

Db 1595294 CAGATCTTTTAGTGAAGGGGGGATTTACAACCCGGAAGATTGACGAGTAAAGAG 1595235  
 Qy 1337 TCTGGGCAACGGTGGGGAAGTGTGGTCTCAACTTTGAAGCGTGTCTCGACGACCA 1396  
 Db 1595234 TTTGGGGCAACGGTGGGGAAGTGTGGTCTCAACTTTGAAGCGTGTCTCGACGACCA 1595175  
 Qy 1397 ACATCATCAAGAGATCCAA 1416  
 Db 1595174 ATGTGATTGAAAATAATTTAA 1595155

## RESULT 8

US-09-252-991A-16413  
 ; Sequence 16413, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: ABRUIGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 16413  
 ; LENGTH: 1581  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 ; US-09-252-991A-16413

Query Match 34.5%; Score 494.2; DB 3; Length 1581;  
 Best Local Similarity 59.8%; Pred. No. 6.9e-134;  
 Matches 848; Conservative 0; Mismatches 568; Indels 3; Gaps 1;

Qy 1 ATGAAATGAACGCTGCCAGATTGAAACGTGACAGAGTGAATGCTGGTGTGATGTATG 60  
 Db 160 ATGAAATGTCATGACCCCTGTTTCGACAGGCCCTGGTGTGATGTGATGTATG 219  
 Qy 61 CTGATGCTTACTGATGAGCGCCCAACGATGATCTCGCCGGAAGCGCGTGCCTG 120  
 Db 220 CTGACCGCTATGATGATGAGCGCACTTCGCGCATTTCCGCGAGGCCGCGTGCCTG 219  
 Qy 121 GTTAAATGAATCATGAAAGAGTCCGCGCGCGCGCTAAAGTGGCGATGATATC 180  
 Db 280 GTCCGCTGCAACAGCAAGAGACCGCCGCGCGCGCGCAACGTCCGCTGAATC 339  
 Qy 181 GTTCTCTCGGTGTATGACGCGCTGTGGGTGACGCGCATGACATGACGCGC 240  
 Db 340 GCGCGCTGAGCGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 399  
 Qy 241 GCGCTGATTAATCTGTGCGCGACGTCAAGTCAATGCGACTTGTCTGTACGACG 300  
 Db 400 AGCTTGGCCAAACGCTCAAGCGCGCTGTGAGTGAACGCGCTTTCAGCGCATGATGC 459  
 Qy 301 CATCCAGCATTAACAATTAACGGTACTTCCGCAACCAACGCTGATCCGTGAT 360  
 Db 460 CAGCGCAATGCTCAAGTGTGATGATGATGATGATGATGATGATGATGATGATG 519  
 Qy 361 TTTGAAAGAGTTTGAAGTGTGATCCGACGCGCTGACGAGCGGATTAATCAGCG 420  
 Db 520 TTTGAGAGACCGTTC--CGACGAGCGCGCGCGCTGTGAGCGTGTGATGATGATG 576  
 Qy 421 CTGAGTTGATGCGCGCTGTGCTTCTGTGATCAACGCAAGTGTGCGTGCAGAGCTA 480  
 Db 577 CTGCGCAAGATGCAAGT 636  
 Qy 481 CAGCAGATGATCAACTGGCGCGTAAAGCGGCTGTCCGCTGTGATGATCAAAAGT 540  
 Db 637 CAGGTGCTGATTCAGCGCGCGCGCGCGCAACATTCGCGTACTGCGCATCCCAAGGC 696

Qy 541 ACCGATTTGAGCGCTACCGCGCGCTACGCTGTAAACCGCAATCTCTCGAATTTGAA 600  
 Db 697 AAGACTTGGCATATGATGAGCGCGCGCTATCATCCCAACCTGTCCGAATTCGAG 756  
 Qy 601 GCTGTGTGCTGATTAATGTAAGCCGAAGAGATTTGAGCGCGCATGAAATGAT 660  
 Db 757 ACCATGTGCGCGCTGTGCGCGCAAGAGCGAATCTGTGCGCAAGGCGCGCTGATG 816  
 Qy 661 GCGGATTAAGATCTTCGCGCTGTGATGATCCCGTTCGGAACGCGTATGTGCTGCTG 720  
 Db 817 AGCAACTGCACTCTGT 876  
 Qy 721 CAACCGGATTAAGCGCGCTGTGATGATGATCCCAAGCGCGAGGATGATGATGATG 780  
 Db 877 GCGATGCGCAAGCGCGCGCTGTGATGATGATGATGATGATGATGATGATGATGATG 936  
 Qy 781 GGTGCGCGCAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840  
 Db 937 GGTGCGCGCAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 996  
 Qy 841 GAAGAACTGCTTCTTTTTCGCAATGCGCGCGCTGTGATGATGATGATGATGATG 900  
 Db 997 CCGTCCGCGCGCTGTGCAACCTGCGCGCGCGCATGATGATGATGATGATGATGATG 1056  
 Qy 901 TCCAGGATTTGCGCGATGATGATGATGATGATGATGATGATGATGATGATGATG 960  
 Db 1057 GCGCGATCAAGCG 1116  
 Qy 961 GCGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1020  
 Db 1117 GCGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1176  
 Qy 1021 AAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1080  
 Db 1177 AAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1236  
 Qy 1081 GCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1140  
 Db 1237 GAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1296  
 Qy 1141 AAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1200  
 Db 1297 ACTGCGTGAAGGCGGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1356  
 Qy 1201 GCGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1260  
 Db 1357 GCGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1416  
 Qy 1261 ATGCGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1320  
 Db 1417 CTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1476  
 Qy 1321 GCGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1380  
 Db 1477 GTCGCGCGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1536  
 Qy 1381 GATTGCTGAGCAACATCATCAAGATGATCAACAG 1419  
 Db 1537 AACAGCTCCACCACTCCCATGTCGAGAAAGATCCGCAAG 1575

## RESULT 9

US-09-252-991A-16304  
 ; Sequence 16304, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: ABRUIGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788



PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 16304  
LENGTH: 1251  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-16304

Query Match 27.3%; Score 391; DB 3; Length 1251;  
Best Local Similarity 60.0%; Pred. No. 8.5e-104;  
Matches 671; Conservative 0; Mismatches 445; Indels 3; Gaps 1;

```
Qy 301 CATCGACCATTTACCAATTACGGGTACTTCCCGCAACCAACGCTGATCCGCTGGAT 360
Db 2 CAGCGACCATCTGACAGCTCGGGTCATGAGTGTCAACGACCACTGCTCGGGGTGAC 61
Qy 361 TTGGAAGAGGTTTGAAGGTGTGATCCGACGCGCTGACGAGCGGATTATCAGGCG 420
Db 62 TTGGAAGAGGTTTGAAGGTGTGATCCGACGCGCGCTGACGAGCTGAGATCGCTG 118
Qy 421 CTGAGTTGATTTGGCGCGCTGCTTTCTGATCAACGCAAGTGGCTGGCAAGCTA 480
Db 119 CTGGCCAAAGTTCAGAGTGTGTGTGTCTGTCCGATACGCGCAAGGCGGCTACAGAACAC 178
Qy 481 CAGCAGATGATCAACTGCGCGGTAAAGCGGGTGTTCGGGTGCTGATTGATCCAAAGGT 540
Db 179 CAGGTGCTGATCAGCGCGCGCGCGCGCAACATTCGGATCTGCGCGGATCCAGGCG 238
Qy 541 ACCGATTTTGAACGCTACCGCGCGCTACGCTGTTAACCGCGGAATCTCTCGGAATTTGAA 600
Db 239 AAGGACTTCGCGCATCTATCGGGCGCGCGCGCTGATCAACCGGAACCTGTCCGAATTCGAG 298
Qy 601 GCTGTTGCTGTAATGTAAGACGGAAGAGATTTGTAACCGCGCGGATTAATCTGATT 660
Db 299 ACCTATCGTCCGCGCTGTCGCGCAAGCCGAACTGCTGCGCAAGGCGCGCGCTGATG 358
Qy 661 GCCGATTTGCAACTCTGGGCTCTGTTAGTGAACCGGTTCCGAACGAGGATATGCTGCTG 720
Db 359 AGCGAATCTGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 418
Qy 721 CAACCGGTTAAAGCGCGCTGCTGATATGCAACCGCAAGCGAGAGTGTATGATGATTAAC 780
Db 419 CGCGATGCGCAAGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 478
Qy 781 GGTGCGCGCGCAACGCTGATTTGGGCTCTGCGCGCAACGCTGCGAGCGGATTAATTCGCTG 840
Db 479 GGTGCGCGCGCAATCGCTCATCTCAACCTGCGCGCGCGCTGCGCGCGCGCGAGAGCTG 538
Qy 841 GAAGAAGCTGCTTCTTGGCAATGCGCGCGCTGCGCGCGCTGCGCGCGCTGCGCGCGCTG 900
Db 539 CCGTCCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 598
Qy 901 TCACGAGTTTCCGCGATGAGCTGGAATAATCTGTACGTGACGTGACATACAGGCTTT 960
Db 599 GCGGCGATCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 658
Qy 961 GCGGCGATGACCGGAAGAACTGAAGTGGCGCGTGAAGCGCGCGCGCGCGCGCGCGCGCG 1020
Db 659 GCGGCGCTGCGCGCTGGAACAAATGCTGTGCGCAATCGAAGAGCGCGCGCGCGCGCGCG 718
Qy 1021 AAGTGTGATGATCAACGAGTGTCTTTGACATCTCGACGCGCGCGCGCGCGCGCGCGCG 1080
Db 719 AAGATGCTTCAACCAATGAGCTGCTTCAATCTTCAACGCGCGCGCGCGCGCGCGCGCG 778
Qy 1081 GCAAAATGCGCGCAAGCTGAGTGAACGCTTGAATTTGGCGCGCGCGCGCGCGCGCGCG 1140
Db 779 GAAAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 838
Qy 1141 AAGAGGCTGAAAGGAGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1200
Db 839 ACTGCGCTGAAAGGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 898
```

```
Qy 1201 GCGCATGTGAAGCGGTCGATCGGTAGTGTGCTTTGAAGAGGACAGCGCGAGCGCTTG 1260
Db 899 GCGGCGCTGCGCGCGGTCGATCGGTGTGTGAGCTTCCCGAAGACATCCCGAGCGCTG 958
Qy 1261 ATGCGCGGATCTTGGCAATCTGCTGTGAAGCGCGCGCGCATTAACGGAAGAGATT 1320
Db 959 CTCGACAGGTGCGTCCGACGCTGTGTCAAGGGCGCGCGATTAACGCGCGTGAAGAGGTG 1018
Qy 1321 GCGGAGATTAAGATCTGCGCGCAACGCGTGGCGGAAGTGTGTGCTCAACTTGAAGAC 1380
Db 1019 GTGCGCGCGCGATGCTCAAGGCTTAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1078
Qy 1381 GATTGCTGACGACCAACATCATCAAGAGATCCACAG 1419
Db 1079 AACAGCTTCAACACCGCGCATGCTGAGAGATCCGCGAG 1117
```

RESULT 10  
US-09-252-991A-16178/c  
Sequence 16178, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 16178  
LENGTH: 963  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-16178

Query Match 24.4%; Score 350.2; DB 3; Length 963;  
Best Local Similarity 60.8%; Pred. No. 6.2e-92;  
Matches 571; Conservative 0; Mismatches 368; Indels 0; Gaps 0;

```
Qy 481 CAGCAGATGATCCAACTGCGCGCGTAAAGCGGTTCCGGTGTGATGATCCAAAGGT 540
Db 951 CAGGTGCTATTCAGAGCGCGCGCGCGCGCGCGCAACATTCGGTATCGGCCATCCAGGCG 892
Qy 541 ACCGATTTTGAAGCGCTACCGCGCGCTACGCTGTTAACCGCGAATCTCTCGAATTTGAA 600
Db 891 AAGGACTTGCACATCTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 832
Qy 601 GCTGTTGCTGTAATGTAAGACCGGAAGAGATTTGTAAGCGCGCGCGCATGAATCGATT 660
Db 831 ACCATGCTGCGCGCTGTCGCGCAAGCGCAACCTGCTGCGCAAGGCGCGCGCTGATG 772
Qy 772 GCCGATTTGCAACTCTGAGCTCTGTTAGTGAACCGGTTCCGAACAGGGTATGTGCTGCTG 720
Db 771 AGGGAATCGACCTCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 712
Qy 721 CAACCGGTTAAAGCGCGCTGATATGCAACCGCAAGCGCGGAAGTGTATGACGTTAAC 780
Db 711 CGGATATGCGACGCGCGCGCTGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 652
Qy 781 GGTGCGCGCGCAACGCTGATTTGGCGCTGCTGCGCGCAACGCTGCGCAACCGGTAATTCGCTG 840
Db 651 GGTGCGCGCGCATGCGTCACTCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 592
Qy 841 GAAGAAGCTGCTTCTTGGCAATGCGCGCGCTGAGTGTGCTGCGCAACTGCGAAGC 900
Db 591 CCGTCCGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 532
Qy 901 TCACGAGTTTCCGCGATGAGCTGGAATAATCTGTACGTGAGAGTGTGCAATACAGGCTTT 960
```



Db 531 GCGGCGATCAGCGCGCCGAACTGCTGCGCGGTGAGCGCGAGCAGGAGGTTCCGAGCGT 472  
Qy 961 GCGGTGATGACCGAAGAGAACTGAGCTGCGCTGAGCGGAGCGGCTTAAAGTGTTGAA 1020  
Db 471 GCGGTGCGCGCTGCGAGCAATGCTGCTGCAATCGAAGAGCGCGCGCCGAGCGAG 412  
Qy 1201 AAGTGTGATGACCAACGCTGCTTGAACATCTGACACGCGCGGCAAGTCTCTAATCTG 1080  
Db 411 AAGATGCTTCAACCAATGCTGCTGCAACATCTTCAACGCGCGCAAGTCACTTACCTC 352  
Qy 1081 GCAATATCCCGCAAGCTGCTGAGACCGCTGATGTTGCTGCTCAACAGCGATGCTTCAAC 1140  
Db 351 GAAACAGCGCGCGCGCGAGCGCGAGCGCTGATGCTGCGGGTCAACAGAGCGCTTCCGTC 292  
Qy 1141 AAGCGCTGAAGAGGAGTTCCCGCGCGTAAACCACTCGAAGAGCGGTATGTTGCTG 1200  
Db 291 ACTGCGCTGAAGGCGGTGCGCGCGCAATCAATCGGTGAGCGCGCGCATGCGGTACTC 232  
Qy 1201 GCGGCACTGGAAGGCGGTGAGCTGCTGCTGTTGAAGAGCAGCGCGCAAGCGCTTG 1260  
Db 231 GCGCGGCTCGCGCGGTGAGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 172  
Qy 1261 ATCGCGCGGATCTTGCAGATCTGCTGCTGGAAGAGCGCGGACTTAAACAGAGAGATT 1320  
Db 171 CTGAGAGAGGTGCGCTCGGAGAGTGTCTGCTGCAAGGCGCGGATTCGCGCTGAGAGGTG 112  
Qy 1321 GCGCGAGTAAAGAGTCTGCGGCGCAAGCTGCGGAGTGTGCTGCAACTTGAAGAC 1380  
Db 111 GTCGCGCGCGAGATGCTCAAGGCTTACGCGCGAGGTACGCGGTGCTGCGCTGCTGAG 52  
Qy 1381 GTTGCTCGACGACCAATCATCAAGAGATCCAGAG 1419  
Db 51 AACAGCTCACACCGCGCATGCTGAAGAGATCCGAG 13

## RESULT 11

US-09-790-988-1/c  
Sequence 1, Application US/09790988  
Patent No. 6632935

GENERAL INFORMATION:  
APPLICANT: SHIGENOBU, SHUJI  
APPLICANT: MATANABE, HIDEMI  
APPLICANT: HATTORI, MASAHIRA  
APPLICANT: SAKAKI, YOSHIYUKI  
TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS  
FILE REFERENCE: 081356/0159  
CURRENT APPLICATION NUMBER: US/09/790,988  
CURRENT FILING DATE: 2001-02-23  
PRIOR APPLICATION NUMBER: JP2000-107160  
PRIOR FILING DATE: 2000-04-07  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 640681  
TYPE: DNA  
ORGANISM: Buchnera sp.  
US-09-790-988-1

Query Match 16.6%; Score 238.4; DB 3; Length 640681;  
Best Local Similarity 53.4%; Pred. No. 4,76-58;  
Matches 500; Conservative 0; Mismatches 436; Indels 0; Gaps 0;

Qy 1 ATGAAGATTAACGCTGCGAGATTGAAAGTGCAGAGTGAATGCTGCTGATGATGATG 60  
Db 66058 ATGAAGATTAACGCTGCGAGATTGAAAGTGCAGAGTGAATGCTGCTGATGATGATG 65999  
Qy 61 CTGAGATGTTACTGTTACGCGCCACAGCTGTTCTCGCGAGAGCGCGGTCCCGCTG 120  
Db 65998 CTGAGATGTTACTGTTACGCGCCACAGCTGTTCTCGCGAGAGCGCGGTCCCGCTG 65939  
Qy 121 GTTAAAGTGAATACATGAGAGAGCTCGCGCGCGCGCTTAAAGTGCAGATGATATATC 180

Db 65938 GTACCAATTAATAAATCAAGAAACAAGTGAAGTCTCAATGATGATTAATAATAT 65879  
Qy 181 GCTTCTCGGTGCTATGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240  
Db 65878 GCAAGATGTTGGGGATTAATAAATGTTGTTTATGGAATGATTAATGAAGATT 65819  
Qy 241 GCGGTGATTAATCTGCGCGAGCTCAAGCTCAATGCAACTGCTGCTGCTGCTGCTG 300  
Db 65818 ATATTAATAAATCTTATGATGATCAATTCGATGATGATGATGATGATGATGAT 65759  
Qy 301 CATCCGACATTAACAATTAACGAGTACTTCCCGCAACCAAGCTGATCCGCTGAT 360  
Db 65758 AATTAATCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 65699  
Qy 361 TTGAAGAGATTTGAGAGTGTGATCGGACCGCTGACAGAGGATTAATCAGCGG 420  
Db 65698 TTTCAGAAAATAATTAATTTCTAATAAACAATCTTATCATCAAAAATAATTAAT 65639  
Qy 421 CTGAGTTGATGCGCGCTGCTGCTTCTGACCAAGGCTGCGCGAGAGCTG 480  
Db 65638 ATATGCTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 65579  
Qy 481 CAGCAGATGATCAACTGCGCGCTGAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 540  
Db 65578 CAATAATTAATGATCTTGCAAAAATAATGCTATCCCTATCTTATGATCTTAAAG 65519  
Qy 541 ACCGATTTGAGGCTACCGCGCGCTGACGCTGTTAACCGGATCTCTGGAATTTGA 600  
Db 65518 ATGATTTTAAATAAATAATTCGAGAGCTGATTTTAAACCAATCTTTTGAATTTGA 65459  
Qy 601 GCTGTTGCTGATTAATGTAAGCAAGAAAGATTTGAGCGCGCATGAATGAT 660  
Db 65458 AAGATGTTGCAAAATTTTAAAGAAATGAATTAATTAATTAATTAATTAATTAAT 65399  
Qy 661 GCGGATTAACAATCTGCGCTGCTGTTAGTGAACCGCTTCCGACAGGATATGCTGCTG 720  
Db 65398 TCTGAATCAAGTATCAAGTATTAATTAATTAATTAATTAATTAATTAATTAAT 65339  
Qy 721 CAACCGGTAAAGCGCGCTGCTATGCAACCAAGCGAGAGATGTTAGCTTACC 780  
Db 65338 CAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 65279  
Qy 781 GGTGCGGCGACAGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840  
Db 65278 GTGCGAGAGATCAAGTATGCTATTAATTAATTAATTAATTAATTAATTAATTA 65219  
Qy 841 GAAGAAGCTGCTTCTTTCGCAATGCGCGCTGCGCTGCTGCTGCTGCTGCTGCTG 900  
Db 65218 GAAGAAGCTGCTTCTTTCGCAATGCGCGCTGCGCTGCTGCTGCTGCTGCTGCTG 65159  
Qy 901 TCAACGCTTTCGCGATCGAGCTGGAATGCTGTA 936  
Db 65158 GAAGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 65123

## RESULT 12

US-09-252-991A-15975/c  
Sequence 15975, Application US/09252991A  
Patent No. 6551795

GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 15975  
LENGTH: 501

TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-15975

Query Match 12.3%; Score 176.2; DB 3; Length 501;  
Best Local Similarity 59.5%; Pred. No. 3e-41;  
Matches 298; Conservative 0; Mismatches 203; Indels 0; Gaps 0;

QY 555 CTACGGGGGGGCTACGGCTGTTAAAGCCGAATCTCGGAATTTAAAGTGTGCGGTA 614  
DB 501 CTATCGGGGGCGCCAGCTGATACCCCGAATCTGGAATTCAGACCATCGTCGGCG 442  
QY 615 ATGTAAAGCCGAGAGAGATTTGTAGCGCGGATGAACTGATTCGATTAAGAACT 674  
DB 441 TTGGCGGAGAGACCGAATCGTGGCAAGGCGCAAGCGCTGATGAGCAACTGACCT 382  
QY 675 CTGGCTCTGTAGTACCCGTTCCGAACAGGATATGCTGCTGCAACCGGGTAAAGC 734  
DB 381 CGGTGCTGTGCTGTCACCCCGGCGAGCATGACATCCCTGCTCGCGATGCGCAGCC 322  
QY 735 GCGGCTGATATGTCACCCGAGGAGGAGAGTATGATGATGATGCGGTCGCGGAC 794  
DB 321 GGGCTGACCTGCGCGGCGCGGCGGAGAGTGTGACGTCAACCGTGCAGGAGTAC 262  
QY 795 GGTGATTGCGCTCTGCGCGGAGCGCTGCGACGCGGATTAATTCGCTGAGAAAGAGCTCT 854  
DB 261 GGTGATCTGCAACCTGCGCGGCGGCTTGGCGCGGAGAGAGCTGCGCGGTGGG 202  
QY 855 CTTTGGCAATCGCGCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 914  
DB 201 CTTGGCCAACTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 142  
QY 915 GATGAGCTGGAATAATGCTGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 974  
DB 141 GCGGAACTGCTGCGCGCGGCTGCGCGGAGCGAGAGAGGTTCCAGGCTGCGCTGGGCT 82  
QY 975 AGAGAACTGAAGTGGCGGCTGAGCGGAGCGCGTAAAGTGTGTAAGAAAGTGTGATGAC 1034  
DB 81 GGAACAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 22  
QY 1035 CAAGGCTGCTTGTGACATCTT 1055  
DB 21 CAATGGCTGCTTGCACATCTT 1

RESULT 13  
US-08-651-155B-185  
Sequence 185, Application US/08651155B  
GENERAL INFORMATION:  
PATENT No. 6365401  
APPLICANT: Mahan Dr., Michael J.  
APPLICANT: Comer Mr., Christopher P.  
TITLE OF INVENTION: METHOD AND PROBES FOR THE IDENTIFICATION  
OF MICROBIAL GENES SPECIFICALLY INDUCED DURING HOST  
INFECTION  
NUMBER OF SEQUENCES: 255  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Christman, Byron & Johnson, P.C.  
STREET: 1900 Fifteenth Street  
CITY: Boulder  
STATE: CO  
COUNTRY: USA  
ZIP: 80302  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/651.155B  
FILING DATE: 17-MAY-1996  
CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:  
NAME: Petersen Mr., Steven C.  
REGISTRATION NUMBER: 36,238  
REFERENCE/DOCKET NUMBER: 17060.1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 303/546-1300  
TELEFAX: 303/449-5426  
TELEX: ABAL475  
INFORMATION FOR SEQ ID NO: 185:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 347 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHEICAL: NO  
ANTI-SENSE: YES  
US-08-651-155B-185

Query Match 10.4%; Score 149.4; DB 3; Length 347;  
Best Local Similarity 74.8%; Pred. No. 1.7e-33;  
Matches 240; Conservative 0; Mismatches 76; Indels 5; Gaps 4;

QY 64 GATGTTACTGTGTAAGCGCCGACAGTCTGATCTCGCCGGAAGCGCGTGCCTGCTT 123  
DB 1 GATGCTATTGTGATGAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60  
QY 124 AAGTGAATTCATGCAAGAAAGTCCGGGCGGCGGCTTAAGTGGGATGAATATGCT 183  
DB 61 AAGTGAATTCATGCAAGAAAGTCCGGGCGGCGGCTTAAGTGGGATGAATATGCT 120  
QY 184 TCTCTGCTGCTATGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 243  
DB 121 TGCTGTGAGGAGCAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 179  
QY 244 CTGAGTAAATCTTGGCCGAGCTGCAACGTCAATG-CGACTTGTGTTCTGATCCAGCA 302  
DB 180 CTGAGCAAAACGCTGCGGAGGTCAATGTGAAGTGCAGACTTGTGTTCTGCTGCAAGCA 239  
QY 303 TCGGACATTAACAATTAACGGGTACTTTC-CGGCAACCAAGCATGATCGTGTGATT 361  
DB 240 TCGGACATTAACAATTAACGGGTACTTTC-CGGCAACCAAGCATGATCGTGTGATT 297  
QY 362 TTGAAGAAGGTTTGAAGGTG 382  
DB 298 TTGAAGAAGGCTTTGAAGATG 318

RESULT 14  
US-09-194-036B-185  
Sequence 185, Application US/09194036B  
PATENT No. 6548246  
GENERAL INFORMATION:  
APPLICANT: Mahan, Michael J.  
APPLICANT: Comer, Christopher P.  
TITLE OF INVENTION: METHOD AND PROBES FOR THE IDENTIFICATION  
OF MICROBIAL GENES SPECIFICALLY INDUCED DURING HOST  
INFECTION  
NUMBER OF SEQUENCES: 255  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Morison & Foerster  
STREET: 755 Page Mill Road  
CITY: Mountain View  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/194, 036B  
FILING DATE: 17-NO. 6548246-1998  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US97/08208  
FILING DATE: 1997-05-16  
APPLICATION NUMBER: US 08/651,155  
FILING DATE: 1996-05-17  
ATTORNEY/AGENT INFORMATION:  
NAME: Shantanu Basu  
REGISTRATION NUMBER: 43,318  
REFERENCE/DOCKET NUMBER: 220002060601  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 813-5995  
TELEFAX: (650) 494-0792  
INFORMATION FOR SEQ ID NO: 185:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 347 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
HYPOTHETICAL: NO  
ANTI-SENSE: YES  
ORIGINAL SOURCE:  
ORGANISM: DNA (other)  
SEQUENCE DESCRIPTION: SEQ ID NO: 185:  
US-09-194-036B-185

Query Match 10.4%; Score 149.4; DB 3; Length 347;  
Best Local Similarity 74.8%; Pred. No. 1.7e-33;  
Matches 240; Conservative 0; Mismatches 76; Indels 5; Gaps 4;

QY 64 GATGTTACTGTAACGCGCCCAACAGTCGATCTCGCGGAGGCGCGGTGCTT 123  
DB 1 GATGCTATTGTTGATGCGCCCACTTGCGTATTTACCGGAGGCGCGGTGCTT 60

QY 124 AAGTGAATACCATGAGAGAGTCCGGGCGCGGCTACGTTGCGGATATATGCT 183  
DB 61 AAGGTAATATCGTTGAGAGAGCGCGGCGCGGAGAGCGTGGCGATTAACATTCG 120

QY 184 TCTCTCGTCTTAATGACGCGCTGTCGCGTTGACGCGCATGACGACGCGCG 243  
DB 121 TGCCTGGAGAGCGAGCGCGTCTGTGCTGCGCTGAACGCG -TTATTGATGAGCGCGCGCGC 179

QY 244 CTGAGTAATATCTGCGCGGACGTCACATGATG -GACTTCTGTTCTGACGAGCA 302  
DB 180 CTGAGCAAAAGCGTGGCGGAGTCAATGTAAGTCCGACTTCTGTGCGGAGCA 239

QY 303 TCCGACCATTCACAAATTAGCGGTACTTC -CCGCAACCAAGCTGATCCGCTGAT 361  
DB 240 TCCGACCATTCACAAAGTCAAGTCACTATCACTAATCAAGCACTCATTCGTTTG -AT 297

QY 362 TTGAAGAAGTTTCGAAGTG 382  
DB 298 TTGAAGAAGCTTTGAGGATG 318

RESULT 15  
US-09-252-991A-16179/c  
Sequence 16179, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 16179  
LENGTH: 789  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-16179

Query Match 10.1%; Score 144.8; DB 3; Length 789;  
Best Local Similarity 58.3%; Pred. No. 5.3e-12;  
Matches 273; Conservative 0; Mismatches 192; Indels 3; Gaps 1;

QY 1 ATGAAGTAAAGCTGCGGAGTTTGAACGTGACAGAGTATGTTGTTGATGATG 60  
DB 465 ATGAAGTATTCATGCGCGCTTTCGACCAAGCCCGGTGTTGATGTTGATGATG 406

QY 61 CTGATGTTACTGTAACGCGCCCAACAGTCGATCTGCGCGGAGCGCGGTGCTG 120  
DB 405 CTGACCGCTATTGTCATGCGGAGTCCGACATTCGCGGAGCGCGGTGCTG 346

QY 121 GTTAAAGTAAATACATGAAAGAGTCCGGGCGCGGCTAACGTCGATGAATATC 180  
DB 345 GTCCGCTCGAAGACAGAGAGACCGCCCGCGCGCGCCCAACGTCGCTGAACATC 286

QY 181 GCTTCTCTCGTCTATGACGCGCTGTCGCGTTGAACGATGACGCGCG 240  
DB 285 GCGCGCTGGCGGAGAGCGCTTGTGTCGCGGTCAACGCGCGGAGCGCGCGAC 226

QY 241 GCGCTGATTAATCTGCGCGGACGTCACATGCAATGCACTTCTTTCTGACGAC 300  
DB 225 AGCTGCGCAACAGCTCAAGGCGCGCTGAGTGAACGCGCTTCCAGCGCATGATAC 166

QY 301 CATCCGACATTCACAAATTAGCGGTACTTCCGCAACCAAGCTGATCCGCTGAT 360  
DB 165 CATCCGACATTCGTAAGCTGCGGTATGATGTCACCAAGCACTGCGCGGTGAC 106

QY 361 TTGAAGAAGTTTCGAAGTGTGATCCGACCGCTGACGAGCGGATTAATCAAGCG 420  
DB 105 TTGAAGAAGCGTTT -CGCACGAGCGCGCGCGCTGCGATGAGCGTGTGCTG 49

QY 421 CTGAGTTGATGCGCGCTGTCGCTTCTGACTGACCAAGTGTG 468  
DB 48 CTGCGCAAGGTCAAGGTGTCGTGCTGTCGACTACGCGCAAGGCGCG 1

Search completed: March 18, 2006, 19:21:03  
Job time : 469 secs

**THIS PAGE LEFT BLANK**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

## OM protein - protein search, using sw model

Run on: March 18, 2006, 06:18:38 ; Search time 24 Seconds  
(without alignments)  
568,880 Million cell updates/sec

Title: US-10-520-820-13

Perfect score: 2393  
Sequence: 1 MKTLPFERBAGVWVGDVW.....PDCSTTNIIKKIQDQKKG 477

## Scoring table:

BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 169630 seqs, 26622889 residues

Total number of hits satisfying chosen parameters: 169630

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

Published Applications AA New:  
1: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/PTC\_NEW\_PUB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2393	100.0	477	US-10-520-820-13	Sequence 13, Appl
2	744.5	31.1	320	US-10-467-657-3254	Sequence 3254, Ap
3	480.5	20.1	345	US-11-096-686-11061	Sequence 11061, A
4	341.5	14.3	169	US-10-467-657-548	Sequence 548, App
5	324.5	13.6	169	US-11-096-686-10264	Sequence 10264, A
6	229.5	9.6	323	US-11-132-864-7	Sequence 7, Appl1
7	183.5	7.7	361	US-11-096-568A-6281	Sequence 6281, Ap
8	183.5	7.7	369	US-11-096-568A-6281	Sequence 6280, Ap
9	181.5	7.6	378	US-11-096-568A-31212	Sequence 31212, A
10	181.5	7.6	379	US-11-096-568A-31211	Sequence 31211, A
11	181.5	7.6	705	US-11-096-568A-31090	Sequence 31090, A
12	181.5	7.6	741	US-11-096-568A-31089	Sequence 31089, A
13	181.5	7.6	750	US-11-096-568A-31088	Sequence 31088, A
14	166	6.9	261	US-11-096-568A-31213	Sequence 31213, A
15	164.5	6.9	212	US-11-096-568A-6282	Sequence 6282, Ap
16	154.5	6.5	302	US-10-793-626-3062	Sequence 3062, Ap
17	153	6.4	307	US-11-087-099-4308	Sequence 4308, Ap
18	152.5	6.4	323	US-11-096-568A-10265	Sequence 10265, A
19	152.5	6.4	359	US-11-096-568A-10265	Sequence 10265, A
20	144.5	5.8	310	US-11-087-099-1760	Sequence 1760, Ap
21	139.5	5.8	323	US-11-087-099-1760	Sequence 1760, Ap
22	132	5.5	316	US-11-087-099-9229	Sequence 9229, Ap
23	130.5	5.5	488	US-11-087-099-303	Sequence 303, Appl
24	128	5.3	7968	US-11-143-980-49	Sequence 49, Appl
25	125.5	5.2	544	US-10-467-657-1020	Sequence 1020, Ap

26	125.5	5.2	545	7	US-11-201-916-19	Sequence 19, Appl
27	123.5	5.2	325	7	US-11-087-099-1705	Sequence 1705, Ap
28	123.5	5.2	333	7	US-11-086-568A-31752	Sequence 31752, A
29	120.5	5.0	401	7	US-11-096-568A-34248	Sequence 34248, A
30	118	4.9	314	7	US-11-096-568A-31753	Sequence 31753, A
31	116.5	4.9	372	7	US-11-096-568A-31152	Sequence 31152, A
32	116.5	4.9	404	7	US-11-096-568A-31151	Sequence 31151, A
33	115	4.8	371	7	US-11-096-568A-34249	Sequence 34249, A
34	112	4.7	1571	7	US-11-052-554A-2	Sequence 2, Appl1
35	110.5	4.6	1121	7	US-11-087-099-8532	Sequence 8532, Ap
36	109	4.6	524	7	US-11-082-389-10	Sequence 10, Appl
37	108.5	4.5	335	7	US-11-087-099-4364	Sequence 4364, Ap
38	108	4.5	633	6	US-10-467-657-6628	Sequence 6628, Ap
39	107.5	4.5	325	7	US-11-087-099-7465	Sequence 7465, Ap
40	107	4.5	319	6	US-10-793-626-786	Sequence 786, Ap
41	107	4.5	352	6	US-10-840-688-22	Sequence 22, Appl
42	107	4.5	3507	7	US-11-075-185-7	Sequence 7, Appl1
43	106.5	4.5	537	6	US-10-467-657-4598	Sequence 4598, Ap
44	106.5	4.5	14130	7	US-11-175-689-9	Sequence 9, Appl1
45	106	4.4	319	7	US-11-087-099-880	Sequence 880, Appl

## ALIGNMENTS

RESULT 1  
US-10-520-820-13  
Sequence 13, Application US/10520820  
Publication No. US2006003393A1  
GENERAL INFORMATION:  
APPLICANT: MUTABILIS S.A.  
TITLE OR INVENTION: Pathogenicity determinants which can be used as targets for develop  
TITLE OR INVENTION: means for preventing and controlling bacterial infections and/or  
TITLE OR INVENTION: dissemination  
FILE REFERENCE: 1621  
CURRENT APPLICATION NUMBER: US/10/520, 820  
CURRENT FILING DATE: 2005-01-07  
PRIOR APPLICATION NUMBER: PCT/EP2003/008209  
PRIOR FILING DATE: 2003-07-09  
PRIOR APPLICATION NUMBER: FR 0208636  
PRIOR FILING DATE: 2002-07-09  
NUMBER OF SEQ ID NOS: 32  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 13  
LENGTH: 477  
TYPE: PRT  
ORGANISM: Escherichia coli  
US-10-520-820-13

Query Match 100.0%; Score 2393; DB 6; Length 477;  
Best Local Similarity 100.0%; Pred. No. 2,5e-164;  
Matches 477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MKTLPFERBAGVWVGDVWLDKRYGPTSRISPEAPVPPVKNVTEERPGANVAMNI 60  
DB 1 MKTLPFERBAGVWVGDVWLDKRYGPTSRISPEAPVPPVKNVTEERPGANVAMNI 60  
QY 61 ASIGANRLVGLTGIDDAARLSKSLADVANKCDPVSPTPTTKRLVSRNOQLRLD 120  
DB 61 ASIGANRLVGLTGIDDAARLSKSLADVANKCDPVSPTPTTKRLVSRNOQLRLD 120  
QY 121 FEESGFEVDQPIHERINOMLSISIGLVSDYKKGALASVOQMIOLARKGVVLIDPKG 180  
DB 121 FEESGFEVDQPIHERINOMLSISIGLVSDYKKGALASVOQMIOLARKGVVLIDPKG 180  
QY 181 TDFERYGATILTPNISEFAAVGKCTEERIEYERGWKLADYELSLVLTREOGMSLL 240  
DB 181 TDFERYGATILTPNISEFAAVGKCTEERIEYERGWKLADYELSLVLTREOGMSLL 240  
QY 241 QPGKAPLHMPFOAQEVVDVDTGADPTVIGVLAATLAAGNSIEBAKCFPANAAGVVGKLG 300  
DB 241 QPGKAPLHMPFOAQEVVDVDTGADPTVIGVLAATLAAGNSIEBAKCFPANAAGVVGKLG 300

Qy	301	STVSPFLEKAAVGRADDTGVTWTEBELKLAFAAARAKGKGVMMNGVFDILHAGHSYL	3666
Db	301	STVSPFLEKAAVGRADDTGVTWTEBELKLAFAAARAKGKGVMMNGVFDILHAGHSYL	3666
Qy	361	ANARKLGDRILIVAVNSDASTKRLKGDSPVPNPLEQRMIVLGLBAVDVWVSFFEDDTQRL	420
Db	361	ANARKLGDRILIVAVNSDASTKRLKGDSPVPNPLEQRMIVLGLBAVDVWVSFFEDDTQRL	420
Qy	421	IAGILPDLLVKGGDYPEREIAKSRVWANGGEVILVLPEDGCGSTNNIIKKIQODKKG	477
Db	421	IAGILPDLLVKGGDYPEREIAKSRVWANGGEVILVLPEDGCGSTNNIIKKIQODKKG	477

```

RESULT 2
US-10-467-657-3254
: Sequence 3254, Application US/10467657
: Publication No. US20050260581A1
: GENERAL INFORMATION:
: APPLICANT: CHIRON SpA
: APPLICANT: FONTANA Maria Rita
: APPLICANT: PIZZA Mariagrazia
: APPLICANT: MASIGNANI Vega
: APPLICANT: MONACI Elisabetta
: TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
: FILE REFERENCE:
: CURRENT APPLICATION NUMBER: US/10/467,657
: CURRENT FILING DATE: 2003-08-11
: PRIOR APPLICATION NUMBER: GB-0103424.8
: PRIOR FILING DATE: 2001-02-12
: NUMBER OF SEQ ID NOS: 9218
: SOFTWARE: SeqwIn99, version 1.04
: SEQ ID NO 3254
: LENGTH: 320
: TYPE: PR1
: ORGANISM: Neisseria gonorrhoeae
US-10-467-657-3254

```

Query Match	31.1%	Score 744.5	DB 6	Length 320
Best Local Similarity	51.0%	Pred. No. 3.3e-46		
Matches 157	Conservative 48	Mismatches 102	Indels 1	Gaps 1
Qy	8	FERAGVMVVGDMVMDLRMYGPTSRISPEARVYVVKNTIEERPGGAAVYANNIAGISGANA	67	
Db	14	FAQAKVLVVGDMVMDLRMYFGVDVSRISPEARVYALIGRIDGAGGAANVARIATLGGGA	73	
Qy	68	RLVGLITGIDDAARALSKSLADNVVKDFVSVPTHTPTTKLRYLSRNQQLIRLDFBGEFG	127	
Db	74	GILSTGTGDEAAALDALMVODGVASLYLMRDQOIAITVVLRYVARNQQLIRLDFESHPR	133	
Qy	128	VDPGRLHRRINALSISGLVYISDYAKGALVAGQOMIGLARAGRPVYLIDPGTDEPERY	187	
Db	134	EYLEDEIKRRYRILPEYDILIFSDYKGGLSHISPDIMAKGVGKTVLIDPGDDYETV	193	
Qy	188	GATLITLPNLSEEEAVVGKCKTEEEIVERGMKLIADVELSALLVTRSEQGSMLDQKAPL	247	
Db	194	GATLITTPNCALKEKGVVGSWKNEGDLTEKQONLRHIDLAVALLTSEBSMTLFFSGE-PI	252	
Qy	248	HMPITQAEVYDYVTGAGDPTVIGVLAATLTAAGNSLEBECFPANAAAGVYVKGKLGSTVSPLE	307	
Db	253	YQPTQAEVYDVDSGAGDVTIAGMGILGALAGCTMPEAMYLANTAAAGVVAKLGTAVCSPAE	312	
Qy	308	LENAYRGR	315	
Db	313	LYEALDQ	320	

RESULT 3  
 US-11-098-686-11061  
 ; Sequence 11061, Application US/11098686  
 ; Publication No. US20060024656A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kapur, Vivek and Gbhardt, Connie J.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES

```

; TITLE OF INVENTION: FROM LAWSONIA INTRACELLULARIS AND METHODS OF USING
; FILE REFERENCE: 09531-128001
; CURRENT APPLICATION NUMBER: US/11/098,686
; PCT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: PCT/US03/31318
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: US 60/416,395
; PRIOR FILING DATE: 2002-10-04
; NUMBER OF SEQ ID NOS: 11433
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO: 11061
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Lawsonia intracellularis
US-11-098-686-11061

Query Match      20.1%; Score 480.5; DB 7; Length 345;
Best Local Similarity 37.0%; Pred. No. 3,3e-27;
Matches 117; Conservative 51; Mismatches 127; Indels 21; Gaps 6

```

[illegible]

```

RESULT 4
US-10-467-657-548
; Sequence 548. Application US/10/467657
; Publication NO. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqwIn99, version 1.04
; SEQ ID NO 548
; LENGTH: 169
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-548

```

Query Match	14.3%;	Score 341.5;	DB 6;	Length 169;
Best Local Similarity	51.8%;	Pred. No. 1.1e-17;		
Matches	71;	Conservative	21;	Mismatches 44;
			Indels	1;
			Gaps	1

QY 342 VMTNGVFDILHAGHVSILANARKLGDRLIVAVNSDASTKRL-KGDSRPVPLEQRMIVL 400  
 DB 32 LVFTNGCPDILHAGHVSILANARKLGDRLIVAVNSDASTKRL-KGDSRPVPLEQRMIVL 91  
 QY 401 GALEAVDVMVFEEDTQORLILAGILPDLLVKGDKPEEISAGSEVANGSEVLVLPED 460  
 DB 92 AALLESVDLVWFEDTQORLILAGILPDLLVKGDKPEEISAGSEVANGSEVLVLPED 151  
 QY 461 GCSTTNIKKIQODKKG 477  
 DB 152 QTSSTTKIAKIRAEKG 168

RESULT 5  
 US-11-098-686-10264  
 ; Sequence 10264, Application US/11098686  
 ; Publication No. US20060024696A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kapur, Vivek and Gebhart, Connie J.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES  
 ; TITLE OF INVENTION: FROM LAMSONIA INTRACELLULARS AND METHODS OF USING  
 ; FILE REFERENCE: 09531-128001  
 ; CURRENT APPLICATION NUMBER: US/11/098, 686  
 ; PRIOR FILING DATE: 2005-04-04  
 ; PRIOR APPLICATION NUMBER: PCT/US03/31318  
 ; PRIOR FILING DATE: 2003-10-01  
 ; PRIOR APPLICATION NUMBER: US 60/416,395  
 ; PRIOR FILING DATE: 2002-10-04  
 ; NUMBER OF SEQ ID NOS: 11433  
 ; SOFTWARE: FASTSEQ for Windows Version 4.0  
 ; SEQ ID NO 10264  
 ; LENGTH: 169  
 ; TYPE: PRT  
 ; ORGANISM: Lawsonia intracellularis  
 ; US-11-098-686-10264

Query Match 13.6%; Score 324.5; DB 7; Length 169;  
 Best Local Similarity 46.1%; Pred. No. 1.9e-16;  
 Matches 65; Conservative 27; Mismatches 48; Indels 1; Gaps 1;  
 QY 336 RKGGEKVTMTNGVFDILHAGHVSILANARKLGDRLIVAVNSDASTKRL-KGDSRPVPLE 394  
 DB 25 KSNKKIIFNNGCVDILHAGHVSILANARKLGDRLIVAVNSDASTKRL-KGDSRPVPLE 84  
 QY 395 QRMIVLGALEAVDVMVFEEDTQORLILAGILPDLLVKGDKPEEISAGSEVANGSEVL 454  
 DB 85 IRAFVLALHLEIDFIFFEEDTQORLILAGILPDLLVKGDKPEEISAGSEVANGSEVL 144  
 QY 455 VLNEDGCTTNIKKIQODK 475  
 DB 145 SLPLEGHSTSGLVOKIRNKK 165

RESULT 6  
 US-11-132-864-7  
 ; Sequence 7, Application US/11132864  
 ; Publication No. US20050289670A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Jintul Shi  
 ; APPLICANT: David Erci  
 ; APPLICANT: Lisa Hagen  
 ; APPLICANT: Hongyu Wang  
 ; TITLE OF INVENTION: Plant Myo-Inositol Kinase  
 ; TITLE OF INVENTION: Polynucleotides and Methods of Use  
 ; FILE REFERENCE: 035718/291638  
 ; CURRENT APPLICATION NUMBER: US/11/132, 864  
 ; PRIOR FILING DATE: 2005-05-19  
 ; PRIOR APPLICATION NUMBER: 60/575,000  
 ; PRIOR FILING DATE: 2004-05-20  
 ; NUMBER OF SEQ ID NOS: 48  
 ; SOFTWARE: FASTSEQ for Windows Version 4.0  
 ; SEQ ID NO 7  
 ; LENGTH: 323

TYPE: PRT  
 ORGANISM: Artificial Sequence  
 FEATURE:  
 OTHER INFORMATION: Pfam consensus for pfkb family  
 US-11-132-864-7

Query Match 9.6%; Score 229.5; DB 7; Length 323;  
 Best Local Similarity 29.2%; Pred. No. 3.1e-09;  
 Matches 102; Conservative 35; Mismatches 127; Indels 85; Gaps 14;  
 QY 13 VMTNGVFDILHAGHVSILANARKLGDRLIVAVNSDASTKRL-KGDSRPVPLEQRMIVL 67  
 DB 2 VVVGANLIDLI-PTVGLPFGELN-ERVVSFEKPGAGANVAVALRGNPSYGVK 57  
 QY 68 RLVLGTGIDDAARALSKSLADVNVKCDPVSVPT- 111  
 DB 58 AFIGKVGDDFEFGFLLKKEGVDDVYKVGEGRTGLALVLPDGGERRITVPRGAN 117  
 QY 112 RNOQLRLDPEBGFEGVD- 159  
 DB 118 ADLTLEBLD-BDLLEADILHSGISLVLLPEPTLEALAE- 160  
 QY 160 VQOMIOLARKAGVPLIDPKGTD- 207  
 DB 161 -AAKAGKISFPDPLADPLMSDEBALVELLEPLADILKPEEELTGLKG 213  
 QY 208 TEEIYERGMKLIADYELSL-LVTRSEQMSLQPGKAPLHPTQAO-EVYDVTGAGDT 265  
 DB 214 EVERBALAHLKILAKAVTGLVVTGAGDGLLVTVTG-GEVHVAVPKVKVDTTGA 272  
 QY 266 -VIGVLAATLAA- 305  
 DB 273 FVAGFLAGLITDSDTQDGLDKEBLRFPANAAALVQKKAISLP 321

RESULT 7  
 US-11-096-568A-6281  
 ; Sequence 6281, Application US/11096568A  
 ; Publication No. US20060048240A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Alexandrov, Nikolai et al.  
 ; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
 ; TITLE OF INVENTION: Theby  
 ; FILE REFERENCE: 2750-1592PUS2  
 ; CURRENT APPLICATION NUMBER: US/11/096, 568A  
 ; PRIOR FILING DATE: 2005-04-01  
 ; NUMBER OF SEQ ID NOS: 34471  
 ; SEQ ID NO 6281  
 ; LENGTH: 361  
 ; TYPE: PRT  
 ; ORGANISM: Glycine max  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (1)-(361)  
 ; OTHER INFORMATION: Ceres Seq. ID no. 14314595  
 ; US-11-096-568A-6281

Query Match 7.7%; Score 183.5; DB 7; Length 361;  
 Best Local Similarity 27.0%; Pred. No. 7.4e-06;  
 Matches 85; Conservative 41; Mismatches 146; Indels 43; Gaps 12;  
 QY 13 VMTNGVFDILHAGHVSILANARKLGDRLIVAVNSDASTKRL-KGDSRPVPLEQRMIVL 66  
 DB 56 VVVVGSANAD-IVYVDRILPGSG- 106  
 QY 67 ARVLGTGIDDAARALSKSLADVNVKCD- 122  
 DB 107 TYVVGVDGDAYKRLVYTAGRGGVRLDNLAVASATGHAVALMDSNGNSIVY- 161  
 QY 123 EGFEVD- 178  
 DB 162 -IGANTLSCWPSLPROHLIDVAQAGIVLQREIRIDAVNA-QVAQAAKMGVPLVLA 217

QY 179 KGTDF----ERYRGATLTPNLSEFEAVG-KCTEEBIEVERGKMLADYELSA--LLVT 231  
 Db 218 GGNMGPRLPQINIVNDLSPMETLRLTGMPTSEFEEIOALKC---HEMGAKQVLVK 274  
 QY 232 RSEOGMSLLOPKARPLHMP- QAOEYVDVTGAGDTVIGVLAATLAAGNSLEBACFPANAA 290  
 Db 275 LGHKGSLPFEGEKTIQQPALIAKTVDTTGAGDTFTAAAFVALVEGSKSKECIRFAAAA 334  
 QY 291 AGVVVGKLGSTVSP 305  
 Db 335 ACLCVQVKASPSMP 349

## RESULT 8

US-11-096-568A-6280  
 ; Sequence 6280, Application US/11096568A  
 ; Publication No. US20060048240A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Alexandrov, Nikolai et al.  
 ; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
 ; TITLE OF INVENTION: Thierby  
 ; FILE REFERENCE: 2750-1592PUS2  
 ; CURRENT APPLICATION NUMBER: US/11/096,568A  
 ; CURRENT FILING DATE: 2005-04-01  
 ; NUMBER OF SEQ ID NOS: 34471  
 ; SEQ ID NO 6280  
 ; LENGTH: 369  
 ; TYPE: PRT  
 ; ORGANISM: Glycine max  
 ; FEATURE:  
 ; NAME/KEY: misc.feature  
 ; LOCATION: (1)..(369)  
 ; OTHER INFORMATION: Ceres Seq. ID no. 14314594  
 US-11-096-568A-6280

Query Match 7.6%; Score 183.5; DB 7; Length 369;  
 Best Local Similarity 27.0%; Pred. No. 7.7e-06;  
 Matches 85; Conservative 41; Mismatches 146; Indels 43; Gaps 12;

QY 13 VNVVGDVMDLRVYWGPTSRISPEAPVNVKNTIEERP-----GGANVAMNITASIGAN 66  
 Db 64 VVVVGSANAD--IYEVDRLEPEG-----ETLAASGOTIAGKGANQATCSAKIAYP 114  
 QY 67 ARVLGLTGIDDAARALSKSLADVNVKCD---FVSVPPTHPTITGLRVLSRNOQLIRLDFE 122  
 Db 115 TYTFGVQVDDAYGGLVYVNRAGRGGRVDNLAIVASATGAHVWVLSNGQNSIYV----- 169  
 QY 123 EGFEQVD---PQPLHERINQALSSIGALVLSDYAKGALASVOQMIQLARKAGVPLIDP 178  
 Db 170 --IGANILSCWPSLPRQHLDLVAQAGVLLQREIPDAVNA--QVAQAKKAGVAVVLLDA 225  
 QY 179 KGTDF----ERYRGATLTPNLSEFEAVG-KCTEEBIEVERGKMLADYELSA--LLVT 231  
 Db 226 GGNMGPRLPQINIVNDLSPMETLRLTGMPTSEFEEIOALKC---HEMGAKQVLVK 282  
 QY 232 RSEOGMSLLOPKARPLHMP- QAOEYVDVTGAGDTVIGVLAATLAAGNSLEBACFPANAA 290  
 Db 283 LGHKGSLPFEGEKTIQQPALIAKTVDTTGAGDTFTAAAFVALVEGSKSKECIRFAAAA 342  
 QY 291 AGVVVGKLGSTVSP 305  
 Db 343 ACLCVQVKASPSMP 357

RESULT 9  
 US-11-096-568A-31212  
 ; Sequence 31212, Application US/11096568A  
 ; Publication No. US20060048240A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Alexandrov, Nikolai et al.  
 ; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
 ; TITLE OF INVENTION: Thierby  
 ; FILE REFERENCE: 2750-1592PUS2

; CURRENT APPLICATION NUMBER: US/11/096,568A  
 ; CURRENT FILING DATE: 2005-04-01  
 ; NUMBER OF SEQ ID NOS: 34471  
 ; SEQ ID NO 31212  
 ; LENGTH: 378  
 ; TYPE: PRT  
 ; ORGANISM: Arabidopsis thaliana  
 ; FEATURE:  
 ; NAME/KEY: misc.feature  
 ; LOCATION: (1)..(378)  
 ; OTHER INFORMATION: Ceres Seq. ID no. 13579029  
 US-11-096-568A-31212

Query Match 7.6%; Score 181.5; DB 7; Length 378;  
 Best Local Similarity 25.6%; Pred. No. 1.1e-05;  
 Matches 81; Conservative 45; Mismatches 142; Indels 49; Gaps 11;

QY 28 PTERISPEAPVVPVVKNT-----IEERP-----GGANVAMNITASIGAN 63  
 Db 59 PKSAVDAHAP-PLVVVGSANADIYVEIERLPKEGETISAKTGOTLAGKGANQACGAKL 117  
 QY 64 GANARLVGLTGIDDAARALSKSLAD--VNVKDFV-SVPTHPT---ITGLRVLSRNOQL 117  
 Db 118 MYPTFVGRLGSDAHGKLIABALGDDGCVHLDYRVSNNBPFGHVVWVLSQDQNSIIT 177  
 QY 118 RLDPEEGFEQVD---PQPLHERINQALSSIGALVLSDYAKGALASVOQMIQLARKAGV 173  
 Db 178 -----VGANMKAMPLEMSDDDEIVNAGVLLQREIPDSIN--QVAKAVKAGV 228  
 QY 174 VLIDPKGTDF---ERYRGATLTPNLSEFEAVGKCTEEBIEVERGKMLADYELSA 229  
 Db 229 VILDVGMDPTPIPNELIDSIDLSPMETLSRLTGMPTFEQISQAVAKCHLGVKQVL 288  
 QY 230 VTRSEOGMSLLOPKARPLHMP- QAOEYVDVTGAGDTVIGVLAATLAAGNSLEBACFPAN 288  
 Db 289 VKLSKGSALFVQGEKPIQOSIIPAAQVDTTGAGDTFTAAAFVALVEGSKSKECIRFAA 348  
 QY 289 AAAGVVVGKLGSTVSP 305  
 Db 349 AAASLCVQVKALPSMP 365

RESULT 10  
 US-11-096-568A-31211  
 ; Sequence 31211, Application US/11096568A  
 ; Publication No. US20060048240A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Alexandrov, Nikolai et al.  
 ; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
 ; TITLE OF INVENTION: Thierby  
 ; FILE REFERENCE: 2750-1592PUS2  
 ; CURRENT APPLICATION NUMBER: US/11/096,568A  
 ; CURRENT FILING DATE: 2005-04-01  
 ; NUMBER OF SEQ ID NOS: 34471  
 ; SEQ ID NO 31211  
 ; LENGTH: 379  
 ; TYPE: PRT  
 ; ORGANISM: Arabidopsis thaliana  
 ; FEATURE:  
 ; NAME/KEY: misc.feature  
 ; LOCATION: (1)..(379)  
 ; OTHER INFORMATION: Ceres Seq. ID no. 13579028  
 US-11-096-568A-31211

Query Match 7.6%; Score 181.5; DB 7; Length 379;  
 Best Local Similarity 25.6%; Pred. No. 1.1e-05;  
 Matches 81; Conservative 45; Mismatches 142; Indels 49; Gaps 11;

QY 28 PTERISPEAPVVPVVKNT-----IEERP-----GGANVAMNITASIGAN 63  
 Db 60 PKSAVDAHAP-PLVVVGSANADIYVEIERLPKEGETISAKTGOTLAGKGANQACGAKL 118  
 QY 64 GANARLVGLTGIDDAARALSKSLAD--VNVKDFV-SVPTHPT---ITGLRVLSRNOQL 117



Db 119 MYPTFYGRIGEDAHGGLIALBALDCCGCHLDYVRVNNPETHRAVVMLOSDGQNSIII 178  
Qy 118 RLDPEEGEGVD---POLPHERINQALSSIGALVLSDYAKALASVQOMIQARAKGVP 173  
Db 179 -----VGGANMKAMPEINSDDELIVRNAGIVLQREIPDSINI--QVAKAVKKGAVP 229  
Qy 174 VLIDPKGTDF---ERRRGATLLTPNLSEPEAVVVGKCTEERIEVERGMKLIADYEISALL 229  
Db 230 VILVDGMDTPPIPNELIDSLIDILSPNETELSRLTGMFTETFEQISQVAKCHKLGKQVL 289  
Qy 230 VTRSEQMSLLOPKAKALHP--TQAEVYDVDTGAGDTVIGVLAATLAAGNSLEBACFPAN 288  
Db 290 VKGSGKSALFIOGKRPDIQSIIIPAQVDTTGAGDTFTAFAVAVVEGKSHBCLAFPA 349  
Qy 289 AAGVNVGKLGSTVSP 305  
Db 350 AAASLCVQVKGALPSMP 366

## RESULT 11

US-11-096-568A-31090  
Sequence 31090, Application US/11096568A  
Publication No. US20060048240A1  
GENERAL INFORMATION:  
APPLICANT: Alexandrov, Nikolai et al.  
TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
FILE REFERENCE: 2750-1592PUS2  
CURRENT APPLICATION NUMBER: US/11/096,568A  
CURRENT FILING DATE: 2005-04-01  
NUMBER OF SEQ ID NOS: 34471  
SEQ ID NO 31090  
LENGTH: 705  
TYPE: PRT  
ORGANISM: Arabidopsis thaliana  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)-(705)  
OTHER INFORMATION: Ceres Seq. ID no. 4987030  
US-11-096-568A-31090

Query Match 7.6%; Score 181.5; DB 7; Length 705;  
Best Local Similarity 25.6%; Pred. No. 2.7e-05;  
Matches 81; Conservative 45; Mismatches 142; Indels 49; Gaps 11;

Qy 28 PTERISPEAPVNVKNT-----IEERP-----GGAANYAMNTASL 63  
Db 386 PKSAVDNAHP-PLVVGSANADIVYEIERLPKEGETISAKTGQTLAGKGANQAAAGAKL 444  
Qy 64 GANARLVGLTGIDDAARALSKSLAD--VWVKCDPV-SVPTHTPT---ITKLRVLSRNOQLI 117  
Db 445 MYPTFYGRIGEDAHGGLIALBALDCCGCHLDYVRVNNPETHRAVVMLOSDGQNSIII 504  
Qy 118 RLDPEEGEGVD---POLPHERINQALSSIGALVLSDYAKALASVQOMIQARAKGVP 173  
Db 505 -----VGGANMKAMPEINSDDELIVRNAGIVLQREIPDSINI--QVAKAVKKGAVP 555  
Qy 174 VLIDPKGTDF---ERRRGATLLTPNLSEPEAVVVGKCTEERIEVERGMKLIADYEISALL 229  
Db 556 VILVDGMDTPPIPNELIDSLIDILSPNETELSRLTGMFTETFEQISQVAKCHKLGKQVL 615  
Qy 230 VTRSEQMSLLOPKAKALHP--TQAEVYDVDTGAGDTVIGVLAATLAAGNSLEBACFPAN 288  
Db 616 VKGSGKSALFIOGKRPDIQSIIIPAQVDTTGAGDTFTAFAVAVVEGKSHBCLAFPA 675  
Qy 289 AAGVNVGKLGSTVSP 305  
Db 676 AAASLCVQVKGALPSMP 692

RESULT 12  
US-11-096-568A-31089

Sequence 31089, Application US/11096568A  
Publication No. US20060048240A1  
GENERAL INFORMATION:  
APPLICANT: Alexandrov, Nikolai et al.  
TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
FILE REFERENCE: 2750-1592PUS2  
CURRENT APPLICATION NUMBER: US/11/096,568A  
CURRENT FILING DATE: 2005-04-01  
NUMBER OF SEQ ID NOS: 34471  
SEQ ID NO 31089  
LENGTH: 741  
TYPE: PRT  
ORGANISM: Arabidopsis thaliana  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)-(741)  
OTHER INFORMATION: Ceres Seq. ID no. 4987029  
US-11-096-568A-31089

Query Match 7.6%; Score 181.5; DB 7; Length 741;  
Best Local Similarity 25.6%; Pred. No. 2.9e-05;  
Matches 81; Conservative 45; Mismatches 142; Indels 49; Gaps 11;

Qy 28 PTERISPEAPVNVKNT-----IEERP-----GGAANYAMNTASL 63  
Db 422 PKSAVDNAHP-PLVVGSANADIVYEIERLPKEGETISAKTGQTLAGKGANQAAAGAKL 480  
Qy 64 GANARLVGLTGIDDAARALSKSLAD--VWVKCDPV-SVPTHTPT---ITKLRVLSRNOQLI 117  
Db 481 MYPTFYGRIGEDAHGGLIALBALDCCGCHLDYVRVNNPETHRAVVMLOSDGQNSIII 540  
Qy 118 RLDPEEGEGVD---POLPHERINQALSSIGALVLSDYAKALASVQOMIQARAKGVP 173  
Db 541 -----VGGANMKAMPEINSDDELIVRNAGIVLQREIPDSINI--QVAKAVKKGAVP 591  
Qy 174 VLIDPKGTDF---ERRRGATLLTPNLSEPEAVVVGKCTEERIEVERGMKLIADYEISALL 229  
Db 592 VILVDGMDTPPIPNELIDSLIDILSPNETELSRLTGMFTETFEQISQVAKCHKLGKQVL 651  
Qy 230 VTRSEQMSLLOPKAKALHP--TQAEVYDVDTGAGDTVIGVLAATLAAGNSLEBACFPAN 288  
Db 652 VKGSGKSALFIOGKRPDIQSIIIPAQVDTTGAGDTFTAFAVAVVEGKSHBCLAFPA 711  
Qy 289 AAGVNVGKLGSTVSP 305  
Db 712 AAASLCVQVKGALPSMP 728

## RESULT 13

US-11-096-568A-31088  
Sequence 31088, Application US/11096568A  
Publication No. US20060048240A1  
GENERAL INFORMATION:  
APPLICANT: Alexandrov, Nikolai et al.  
TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
FILE REFERENCE: 2750-1592PUS2  
CURRENT APPLICATION NUMBER: US/11/096,568A  
CURRENT FILING DATE: 2005-04-01  
NUMBER OF SEQ ID NOS: 34471  
SEQ ID NO 31088  
LENGTH: 750  
TYPE: PRT  
ORGANISM: Arabidopsis thaliana  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)-(750)  
OTHER INFORMATION: Ceres Seq. ID no. 4987028  
US-11-096-568A-31088

Query Match 7.6%; Score 181.5; DB 7; Length 750;  
Best Local Similarity 25.6%; Pred. No. 3e-05;

Matches 81; Conservative 45; Mismatches 142; Indels 49; Gaps 11;

```

QY 28 PLSRISSEAPVYVYKVT-----IEERP-----GGANVAMNTASL 63
DB 431 PKAVNDHAP-ELVVGSANADIYERLPKSEETTSAKTGQTLAGGKAMQACAKL 489
QY 64 GANARVLGTGIDDAARALSKSLAD--VNYKCDPV-SVPTHPT---ITKLRVLSRNOOL 117
DB 490 MYPLFYVGRIGEDAHGKLIABALDGCYHLDYRSVNNPEPTGHAVVLMQSDQNSI 549
QY 118 RLDFEEGPEVD---PQLHERINQALSSIGALVSDYAKGALASVQOMIQLARKGV 173
DB 550 -----VGGANMKAMPEISDDLEIVRNAGIYLLQREIPDSINT--QVAKAVKKGVP 600
QY 174 VLIDPKGTF-----ERRRGATLLTPNLSEFEAVVGGCKTEBEIYERGMKLIADYEL 229
DB 601 VILVGGMDPIFENELSDIDLSFNTELSRLTGMPTTFEQISQVAKCHKLGVRQVL 660
QY 230 VTRSEOGMSLLQPGKAPLHMP-TOAEVYDVTGAGDTVIGVLAATLAAGNSLEBACFPAN 288
DB 661 VKGSKGSALFIQGEKPIQOSIIPAAQVDTTGAGDTFTAAFAVAVWGKSHBCLAFPA 720
QY 289 AAAGVYVGGKGTSTVSP 305
DB 721 AAASLCVQVKGALPSMP 737

```

## RESULT 14

```

US-11-096-568A-31213
; Sequence 31213, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096, 568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 31213
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(261)
; OTHER INFORMATION: Ceres Seq. ID no. 13579030
US-11-096-568A-31213

```

Query Match 6.9%; Score 166; DB 7; Length 261;

Best Local Similarity 25.9%; Pred. No. 8.4e-05;

Matches 65; Conservative 42; Mismatches 120; Indels 24; Gaps 8;

```

QY 70 VGLTGIDDAARALSKSLAD--VNYKCDPV-SVPTHPT---ITKLRVLSRNOOLRLDP 123
DB 7 VGRIGEDAHGKLIABALDGCYHLDYRSVNNPEPTGHAVVLMQSDQNSI----- 60
QY 124 GFEGVD---PQLHERINQALSSIGALVSDYAKGALASVQOMIQLARKGVPLIDPK 179
DB 61 -VGGANMKAMPEISDDLEIVRNAGIYLLQREIPDSINT--QVAKAVKKGAVPVLIDVG 117
QY 180 GTDF-----ERRRGATLLTPNLSEFEAVVGGCKTEBEIYERGMKLIADYEL 235
DB 118 GMDPIFENELSDIDLSFNTELSRLTGMPTTFEQISQVAKCHKLGVRQVLVKGSK 177
QY 236 GMSLLQPGKAPLHMP-TOAEVYDVTGAGDTVIGVLAATLAAGNSLEBACFPANAAAGV 294
DB 178 GSALFIQGEKPIQOSIIPAAQVDTTGAGDTFTAAFAVAVWGKSHBCLAFPA 237
QY 295 VGGKGTSTVSP 305
DB 238 VQVKGALPSMP 248

```

## RESULT 15

```

US-11-096-568A-6282
; Sequence 6282, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096, 568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 6282
; LENGTH: 212
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(212)
; OTHER INFORMATION: Ceres Seq. ID no. 14314596
US-11-096-568A-6282

```

Query Match 6.9%; Score 164.5; DB 7; Length 212;

Best Local Similarity 33.6%; Pred. No. 7.9e-05;

Matches 51; Conservative 23; Mismatches 67; Indels 11; Gaps 5;

```

QY 162 QMIQLARKGVPLIDPKGTF-----ERRRGATLLTPNLSEFEAVVGGCKTEBEIYER 216
DB 52 QVAMAKKAGVPLVDAGGMDGPLPPQILNFDILSPNETELARLTGMPRESFEIQA 111
QY 217 MKLIADYELSA--LVTRSEOGMSLLQPGKAPLHMP-TOAEVYDVTGAGDTVIGVLA 273
DB 112 LKC--HEMGAQGVLVKLGHGKGSALFVEGEKTIQAPAILAKTVVDTTGAGDTFTAAFAVA 168
QY 274 LAAGNSLEBACFPANAAAGVYVGGKGTSTVSP 305
DB 169 LVGSKSKBCLAFPA 200

```

Search completed: March 18, 2006, 06:21:47

Job time : 25 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

## OM protein - protein search, using SW model

Run on: March 18, 2006, 06:17:53 ; Search time 168 Seconds

(Without alignments)  
1186.337 Million cell updates/sec

Title: US-10-520-820-13

Perfect score: 2393

Sequence: 1 MKTLPFERAGVWVGDVW.....FEDGCTTNIKKIQDDKKG 477

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA Main:\*

1: /cgn2\_6/ptodate/1/pubpaa/US07\_PUBCOMB.pdp:\*

2: /cgn2\_6/ptodate/1/pubpaa/US08\_PUBCOMB.pdp:\*

3: /cgn2\_6/ptodate/1/pubpaa/US09\_PUBCOMB.pdp:\*

4: /cgn2\_6/ptodate/1/pubpaa/US10A\_PUBCOMB.pdp:\*

5: /cgn2\_6/ptodate/1/pubpaa/US10B\_PUBCOMB.pdp:\*

6: /cgn2\_6/ptodate/1/pubpaa/US11\_PUBCOMB.pdp:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2393	100.0	477	3	US-09-741-669-469 Sequence 469, App
2	2393	100.0	477	3	US-09-912-020-325 Sequence 325, App
3	2393	100.0	477	3	US-09-815-242-10301 Sequence 10301, A
4	2393	100.0	477	4	US-10-282-122A-42614 Sequence 42614, A
5	2393	100.0	477	5	US-10-771-241-325 Sequence 325, App
6	2294	95.9	476	4	US-10-282-122A-56202 Sequence 56202, A
7	2265	94.7	477	3	US-09-815-242-13759 Sequence 13759, A
8	2265	94.7	477	4	US-10-282-122A-75478 Sequence 75478, A
9	2227	93.1	477	4	US-10-282-122A-60277 Sequence 60277, A
10	2196	91.8	463	4	US-10-282-122A-73168 Sequence 73168, A
11	2076	86.8	476	4	US-10-282-122A-78165 Sequence 78165, A
12	2007	83.9	474	4	US-10-282-122A-68720 Sequence 68720, A
13	1712	71.5	476	4	US-10-282-122A-58533 Sequence 58533, A
14	1701	71.1	476	4	US-10-282-122A-67124 Sequence 67124, A
15	1357.5	56.7	474	4	US-10-282-122A-69312 Sequence 69312, A
16	1350.5	56.4	474	3	US-09-815-242-12079 Sequence 12079, A
17	1350.5	56.4	474	4	US-10-282-122A-66769 Sequence 66769, A
18	1338.5	55.9	469	4	US-10-282-122A-67849 Sequence 67849, A
19	863	36.1	461	4	US-10-282-122A-54522 Sequence 54522, A
20	825	34.5	334	4	US-10-282-122A-49479 Sequence 49479, A
21	809.5	33.8	461	4	US-10-282-122A-58820 Sequence 58820, A
22	774.5	32.4	447	4	US-10-335-977-5564 Sequence 5564, App
23	745.5	31.2	323	4	US-10-282-122A-65852 Sequence 65852, A
24	745.5	31.2	320	4	US-10-275-026A-52 Sequence 52, App1
25	744.5	31.1	320	4	US-10-282-122A-65224 Sequence 65224, A
26	734.5	30.7	424	4	US-10-335-977-5563 Sequence 5563, App
27	732.5	30.6	313	4	US-10-282-122A-50827 Sequence 50827, A

28	691.5	28.9	316	4	US-10-282-122A-47619 Sequence 47619, A
29	679.5	28.4	328	4	US-10-282-122A-50163 Sequence 50163, A
30	564.5	23.6	315	4	US-10-369-493-37 Sequence 37, App1
31	420	17.6	315	4	US-10-369-493-10214 Sequence 10214, A
32	362	15.1	161	4	US-10-282-122A-50471 Sequence 50471, A
33	350	14.6	161	4	US-10-282-122A-48155 Sequence 48155, A
34	344	14.4	178	4	US-10-282-122A-49730 Sequence 49730, A
35	341.5	14.3	168	4	US-10-282-122A-65242 Sequence 65242, A
36	340.5	14.2	168	4	US-10-282-122A-65689 Sequence 65689, A
37	332	13.9	162	4	US-10-282-122A-50947 Sequence 50947, A
38	275	11.5	136	4	US-10-282-122A-62085 Sequence 62085, A
39	254.5	10.6	293	4	US-10-369-493-16604 Sequence 16604, A
40	218	9.1	302	5	US-10-501-282-3560 Sequence 3560, App
41	218	9.1	322	5	US-10-501-282-3562 Sequence 3562, App
42	211	8.8	362	4	US-10-424-599-175504 Sequence 175504, A
43	201.5	8.4	319	4	US-10-369-493-17260 Sequence 17260, A
44	199.5	8.3	300	4	US-10-369-493-13944 Sequence 13944, A
45	197	8.2	308	3	US-09-815-242-11854 Sequence 11854, A

## ALIGNMENTS

RESULT 1	
US-09-741-669-469	
Sequence 469, Application US/09741669	
Patent No. US2002022718A1	
GENERAL INFORMATION:	
APPLICANT: Forsyth, R. Allyn	
APPLICANT: Ohlsen, Karl L.	
APPLICANT: Zwick, Judith W.	
TITLE OR INVENTION: Genes identified as required for	
proliferation of B. coli	
FILE REFERENCES: ELITRA.009A	
CURRENT APPLICATION NUMBER: US/09/741,669	
CURRENT FILING DATE: 2000-12-19	
PRIOR APPLICATION NUMBER: US 60/173005	
PRIOR FILING DATE: 1999-12-23	
NUMBER OF SEQ ID NOS: 481	
SOFTWARE: FASTSEQ for Windows Version 4.0	
SEQ ID NO 469	
LENGTH: 477	
TYPE: PRT	
ORGANISM: Escherichia coli	
US-09-741-669-469	
Query Match	100.0%; Score 2393; DB 3; Length 477;
Best Local Similarity	100.0%; Pred. No. 1.3e-198;
Matches 477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 MKTLPFERAGVWVGDVWLDKRYTGPTSRISPEAPVPPVKKVNTTBERGGANVAMNI 60
DB	1 MKTLPFERAGVWVGDVWLDKRYTGPTSRISPEAPVPPVKKVNTTBERGGANVAMNI 60
QY	61 ASIGANARIVGLGIDDAARALSKSLADVVKCDPVPVPHPTITKLRLVSRNOQLRLD 120
DB	61 ASIGANARIVGLGIDDAARALSKSLADVVKCDPVPVPHPTITKLRLVSRNOQLRLD 120
QY	121 FEEGFEVDPPPLHRRINQALSSIGALVLDVAKGALASVOQMIOLARKAGVPLIDPKG 180
DB	121 FEEGFEVDPPPLHRRINQALSSIGALVLDVAKGALASVOQMIOLARKAGVPLIDPKG 180
QY	181 TDERRRGATLTLPNISSEFAVVGKCTBERIYERGGKLIADYELSLVLRBOGMSLL 240
DB	181 TDERRRGATLTLPNISSEFAVVGKCTBERIYERGGKLIADYELSLVLRBOGMSLL 240
QY	241 QPGKAPLHMPYTOAOEYVDVAGADTVYGVLAATLAAGNSIEBACFPANAAAGVVGKLG 300
DB	241 QPGKAPLHMPYTOAOEYVDVAGADTVYGVLAATLAAGNSIEBACFPANAAAGVVGKLG 300
QY	301 STVSPLELNAVAGRADTGVGTBEELKLAVAARRGKGVMTNGVFDILHAGVSYL 360
DB	301 STVSPLELNAVAGRADTGVGTBEELKLAVAARRGKGVMTNGVFDILHAGVSYL 360

Qy 361 ANARKDRLIVAVNSDASTKRLKGDSPVNPPLBORMIVLGALEAVDWPVVSFEEDTPORL 420  
Db 361 ANARKDRLIVAVNSDASTKRLKGDSPVNPPLBORMIVLGALEAVDWPVVSFEEDTPORL 420

Qy 421 IAGILPDLVLKGGDYKPEEIASKEVWANGGEVLVNFEDGCSSTNNIKKIQQDKG 477  
Db 421 IAGILPDLVLKGGDYKPEEIASKEVWANGGEVLVNFEDGCSSTNNIKKIQQDKG 477

## RESULT 2

US-09-912-020-325  
Sequence 325, Application US/09912020  
Patent No. US20020045592A1  
GENERAL INFORMATION:  
APPLICANT: Zykend, Judith  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Trawick, John  
APPLICANT: Forsyth, R. Allen  
APPLICANT: Froelich, Jamie M.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN  
FILE REFERENCE: ELITRA.001DV1  
CURRENT APPLICATION NUMBER: US/09/912,020  
PRIOR FILING DATE: 2001-07-23  
PRIOR APPLICATION NUMBER: 09/492,709  
PRIOR FILING DATE: 2000-01-27  
PRIOR APPLICATION NUMBER: 60/117,405  
PRIOR FILING DATE: 1999-01-27  
NUMBER OF SEQ ID NOS: 485  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 325  
LENGTH: 477  
TYPE: PRT  
ORGANISM: E. Coli  
US-09-912-020-325

Query Match 100.0%; Score 2393; DB 3; Length 477;  
Best Local Similarity 100.0%; Pred. No. 1.3e-198;  
Matches 477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKVTLPEFERAGVWVGDVMDLDRVYGPSTRISPEAPVPVVKVNTIEERPGANVAMNI 60  
Db 1 MKVTLPEFERAGVWVGDVMDLDRVYGPSTRISPEAPVPVVKVNTIEERPGANVAMNI 60  
Qy 61 ASLGANARLVGLTGIDDAARALSKSLADVNCDFVSPHTTTKRLVLSRNOQLRLD 120  
Db 61 ASLGANARLVGLTGIDDAARALSKSLADVNCDFVSPHTTTKRLVLSRNOQLRLD 120  
Qy 121 FEEGEGVDPOPHERINQALSSIGALVLSDYAKGALASVOQMQLARKAGVPLIDPKG 180  
Db 121 FEEGEGVDPOPHERINQALSSIGALVLSDYAKGALASVOQMQLARKAGVPLIDPKG 180  
Qy 181 TDPERYRGATLTPNLSEFAVVGCKTBEIEVERGMKLIADYELSLALVTRSEQMSLL 240  
Db 181 TDPERYRGATLTPNLSEFAVVGCKTBEIEVERGMKLIADYELSLALVTRSEQMSLL 240  
Qy 241 QPGKAPLHPTQAEVYDVDTGAGDTVIGVLAATLAAGNSLEBACFPANAAAGVVGKLG 300  
Db 241 QPGKAPLHPTQAEVYDVDTGAGDTVIGVLAATLAAGNSLEBACFPANAAAGVVGKLG 300  
Qy 301 STVSPIELENAVRGADTGFVMTBEELKLAVAAARKGKRVMTNGVFDILAHGHSYL 360  
Db 301 STVSPIELENAVRGADTGFVMTBEELKLAVAAARKGKRVMTNGVFDILAHGHSYL 360  
Qy 361 ANARKDRLIVAVNSDASTKRLKGDSPVNPPLBORMIVLGALEAVDWPVVSFEEDTPORL 420  
Db 361 ANARKDRLIVAVNSDASTKRLKGDSPVNPPLBORMIVLGALEAVDWPVVSFEEDTPORL 420  
Qy 421 IAGILPDLVLKGGDYKPEEIASKEVWANGGEVLVNFEDGCSSTNNIKKIQQDKG 477

Db 421 IAGILPDLVLKGGDYKPEEIASKEVWANGGEVLVNFEDGCSSTNNIKKIQQDKG 477

## RESULT 3

US-09-815-242-10301  
Sequence 10301, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zykend, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
PRIOR FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 1410  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 10301  
LENGTH: 477  
TYPE: PRT  
ORGANISM: Escherichia coli  
US-09-815-242-10301

Query Match 100.0%; Score 2393; DB 3; Length 477;  
Best Local Similarity 100.0%; Pred. No. 1.3e-198;  
Matches 477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKVTLPEFERAGVWVGDVMDLDRVYGPSTRISPEAPVPVVKVNTIEERPGANVAMNI 60  
Db 1 MKVTLPEFERAGVWVGDVMDLDRVYGPSTRISPEAPVPVVKVNTIEERPGANVAMNI 60  
Qy 61 ASLGANARLVGLTGIDDAARALSKSLADVNCDFVSPHTTTKRLVLSRNOQLRLD 120  
Db 61 ASLGANARLVGLTGIDDAARALSKSLADVNCDFVSPHTTTKRLVLSRNOQLRLD 120  
Qy 121 FEEGEGVDPOPHERINQALSSIGALVLSDYAKGALASVOQMQLARKAGVPLIDPKG 180  
Db 121 FEEGEGVDPOPHERINQALSSIGALVLSDYAKGALASVOQMQLARKAGVPLIDPKG 180  
Qy 181 TDPERYRGATLTPNLSEFAVVGCKTBEIEVERGMKLIADYELSLALVTRSEQMSLL 240  
Db 181 TDPERYRGATLTPNLSEFAVVGCKTBEIEVERGMKLIADYELSLALVTRSEQMSLL 240  
Qy 241 QPGKAPLHPTQAEVYDVDTGAGDTVIGVLAATLAAGNSLEBACFPANAAAGVVGKLG 300  
Db 241 QPGKAPLHPTQAEVYDVDTGAGDTVIGVLAATLAAGNSLEBACFPANAAAGVVGKLG 300  
Qy 301 STVSPIELENAVRGADTGFVMTBEELKLAVAAARKGKRVMTNGVFDILAHGHSYL 360  
Db 301 STVSPIELENAVRGADTGFVMTBEELKLAVAAARKGKRVMTNGVFDILAHGHSYL 360  
Qy 361 ANARKDRLIVAVNSDASTKRLKGDSPVNPPLBORMIVLGALEAVDWPVVSFEEDTPORL 420

Db 361 ANARKDRLIVAVNSDASTKRLKGSRPVNPLEQRMIVGALAEVDMVVSFEEDTPORL 420  
Qy 421 IAGILPDLVKGSDYKPEBIAIGSKYVWANGGEVLVNFEDGCSSTNNIIKKIQODKKG 477  
Db 421 IAGILPDLVKGSDYKPEBIAIGSKYVWANGGEVLVNFEDGCSSTNNIIKKIQODKKG 477

## RESULT 4

US-10-282-122A-42614  
Sequence 42614, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 42614  
LENGTH: 477  
TYPE: PRT  
ORGANISM: Escherichia coli  
US-10-282-122A-42614

Query Match 100.0%; Score 2393; DB 4; Length 477;  
Best Local Similarity 100.0%; Pred. No. 1.3e-198;  
Matches 477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKTTLPEFERAGVWVGDVMDLDRWYGTPTSRISPEAPVPVKNVTIEERPGGAANYAMNI 60  
Db 1 MKTTLPEFERAGVWVGDVMDLDRWYGTPTSRISPEAPVPVKNVTIEERPGGAANYAMNI 60  
Qy 61 ASIGANARLVGLTGIDDAARALSKSLADVNVKCFVSPVPTTTKRLVLSRNOQLRLD 120  
Db 61 ASIGANARLVGLTGIDDAARALSKSLADVNVKCFVSPVPTTTKRLVLSRNOQLRLD 120  
Qy 121 FEEGFEVDPOPLHERINQALSSIGALVLSDYAKGALASVOQMIQLARKAGVPLIDPKG 180  
Db 121 FEEGFEVDPOPLHERINQALSSIGALVLSDYAKGALASVOQMIQLARKAGVPLIDPKG 180

Qy 181 TDEERYGATLTPNLSPEPAVYKCKTEEBIYERGMKLIADVELSALVTRSEOGMSLL 240  
Db 181 TDEERYGATLTPNLSPEPAVYKCKTEEBIYERGMKLIADVELSALVTRSEOGMSLL 240  
Qy 241 QPKAPLHPTQAEVYDVYAGDVIYGVLAATLAAGNSLEBAFCFPANAAAGVVGKLTGT 300  
Db 241 QPKAPLHPTQAEVYDVYAGDVIYGVLAATLAAGNSLEBAFCFPANAAAGVVGKLTGT 300  
Qy 301 STVSPLEENAVAGRADTGFVMTTEBELKLAVAARKGKVMVTNGVFDILHAGVSYL 360  
Db 301 STVSPLEENAVAGRADTGFVMTTEBELKLAVAARKGKVMVTNGVFDILHAGVSYL 360  
Qy 361 ANARKDRLIVAVNSDASTKRLKGSRPVNPLEQRMIVGALAEVDMVVSFEEDTPORL 420  
Db 361 ANARKDRLIVAVNSDASTKRLKGSRPVNPLEQRMIVGALAEVDMVVSFEEDTPORL 420  
Qy 421 IAGILPDLVKGSDYKPEBIAIGSKYVWANGGEVLVNFEDGCSSTNNIIKKIQODKKG 477  
Db 421 IAGILPDLVKGSDYKPEBIAIGSKYVWANGGEVLVNFEDGCSSTNNIIKKIQODKKG 477

## RESULT 5

US-10-771-241-325  
Sequence 325, Application US/10771241  
Publication No. US2004024175A1  
GENERAL INFORMATION:  
APPLICANT: Zyskind, Judith  
APPLICANT: Forsyth, R. Allyn  
TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN  
TITLE OF INVENTION: ESCHERICHIA COLI  
FILE REFERENCE: ELITRA.001C1  
CURRENT APPLICATION NUMBER: US/10/771,241  
CURRENT FILING DATE: 2004-02-03  
PRIOR APPLICATION NUMBER: 09/492,709  
PRIOR FILING DATE: 2000-01-27  
PRIOR APPLICATION NUMBER: 60/117,405  
PRIOR FILING DATE: 1999-01-27  
NUMBER OF SEQ ID NOS: 485  
SOFTWARE: PatSeq for Windows Version 3.0  
SEQ ID NO 325  
LENGTH: 477  
TYPE: PRT  
ORGANISM: E. coli  
US-10-771-241-325

Query Match 100.0%; Score 2393; DB 5; Length 477;  
Best Local Similarity 100.0%; Pred. No. 1.3e-198;  
Matches 477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKTTLPEFERAGVWVGDVMDLDRWYGTPTSRISPEAPVPVKNVTIEERPGGAANYAMNI 60  
Db 1 MKTTLPEFERAGVWVGDVMDLDRWYGTPTSRISPEAPVPVKNVTIEERPGGAANYAMNI 60  
Qy 61 ASIGANARLVGLTGIDDAARALSKSLADVNVKCFVSPVPTTTKRLVLSRNOQLRLD 120  
Db 61 ASIGANARLVGLTGIDDAARALSKSLADVNVKCFVSPVPTTTKRLVLSRNOQLRLD 120  
Qy 121 FEEGFEVDPOPLHERINQALSSIGALVLSDYAKGALASVOQMIQLARKAGVPLIDPKG 180  
Db 121 FEEGFEVDPOPLHERINQALSSIGALVLSDYAKGALASVOQMIQLARKAGVPLIDPKG 180  
Qy 181 TDEERYGATLTPNLSPEPAVYKCKTEEBIYERGMKLIADVELSALVTRSEOGMSLL 240  
Db 181 TDEERYGATLTPNLSPEPAVYKCKTEEBIYERGMKLIADVELSALVTRSEOGMSLL 240  
Qy 241 QPKAPLHPTQAEVYDVYAGDVIYGVLAATLAAGNSLEBAFCFPANAAAGVVGKLTGT 300  
Db 241 QPKAPLHPTQAEVYDVYAGDVIYGVLAATLAAGNSLEBAFCFPANAAAGVVGKLTGT 300  
Qy 301 STVSPLEENAVAGRADTGFVMTTEBELKLAVAARKGKVMVTNGVFDILHAGVSYL 360  
Db 301 STVSPLEENAVAGRADTGFVMTTEBELKLAVAARKGKVMVTNGVFDILHAGVSYL 360

Qy 361 ANARKGDRILIVAVNSDASTKRLKGDSPVNPLEORMIVLGALEAVDWMVSFEEDTPORL 420  
Db 361 ANARKGDRILIVAVNSDASTKRLKGDSPVNPLEORMIVLGALEAVDWMVSFEEDTPORL 420  
Qy 421 IAGILPDLVKGSDYKPEBIAGSKEVWANGGEVLVNFEDGCSSTNIKKIQODKKG 477  
Db 421 IAGILPDLVKGSDYKPEBIAGSKEVWANGGEVLVNFEDGCSSTNIKKIQODKKG 477

RESULT 6  
US-10-282-122A-56202  
Sequence 56202, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT FILING DATE: 2003-02-20  
PRIOR FILING DATE: 2000-03-21  
PRIOR FILING DATE: 2000-03-21  
PRIOR FILING DATE: 2000-05-23  
PRIOR FILING DATE: 2000-05-23  
PRIOR FILING DATE: 2000-05-26  
PRIOR FILING DATE: 2000-09-06  
PRIOR FILING DATE: 2000-09-06  
PRIOR FILING DATE: 2000-09-09  
PRIOR FILING DATE: 2000-10-23  
PRIOR FILING DATE: 2000-11-27  
PRIOR FILING DATE: 2000-12-22  
PRIOR FILING DATE: 2001-02-09  
PRIOR FILING DATE: 2001-02-09  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: Patent version 3.1  
SEQ ID NO 56202  
LENGTH: 476  
TYPE: PRT  
ORGANISM: Enterobacter cloacae  
US-10-282-122A-56202

Query Match 95.9%; Score 2294; DB 4; Length 476;  
Best Local Similarity 94.7%; Pred. No. 5.1e-190;  
Matches 451; Conservative 19; Mismatches 6; Indels 0; Gaps 0;

Qy 1 MKVTLPEERAGVWVGVMDLDRYWGPTSRISPEAPVPVVKNTTIERPGGANVAMNI 60  
Db 1 MKVTLPEERAGVWVGVMDLDRYWGPTSRISPEAPVPVVKNTTIERPGGANVAMNI 60  
Qy 61 ASLGNARLVGLTGIDDAARALSKSLADVNVKDFVSVPTHTTTKRLVLSRNQOLRLD 120  
Db 61 ASLGNARLVGLTGIDDAARALSKSLADVNVKDFVSVPTHTTTKRLVLSRNQOLRLD 120  
Qy 121 FEBEGVDVPOPLHERINQALSIGAVLSDYAKGALASVOOMIQARAKGVPLIDPKG 180

Db 121 FEBEGVDVPEPLHERINQALNIGALVLSDYAKGALASVKTMIQARAKANVPLIDPKG 180  
Qy 181 TDFERRRGATLTLPNISPEPAVVGKCTEBEIVBERMKLIADYBSALVYTRSEOGMSLL 240  
Db 181 TDFERRRGATLTLPNISPEPAVVGKCTEBEIVBERMKLIADYBSALVYTRSEOGMSLL 240  
Qy 241 QPGKAPLHPPTQAEVYDVTGAGDTVIGVLAATLAAGNSLSEBACFPANAAAGVVGKLTGT 300  
Db 241 QPGKAPLHPPTQAEVYDVTGAGDTVIGVLAATLAAGNSLSEBACFPANAAAGVVGKLTGT 300  
Qy 301 STVSPLEENAVGRADTGFVMTTEELKIAVAARRGKGVMTNGVDPDILHAGVSYL 360  
Db 301 STVSPLEENAVGRADTGFVMTTEELKIAVAARRGKGVMTNGVDPDILHAGVSYL 360  
Qy 361 ANARKGDRILIVAVNSDASTKRLKGDSPVNPLEORMIVLGALEAVDWMVSFEEDTPORL 420  
Db 361 ANARKGDRILIVAVNSDASTKRLKGDSPVNPLEORMIVLGALEAVDWMVSFEEDTPORL 420  
Qy 421 IAGILPDLVKGSDYKPEBIAGSKEVWANGGEVLVNFEDGCSSTNIKKIQODKKG 476  
Db 421 IAGILPDLVKGSDYKPEBIAGSKEVWANGGEVLVNFEDGCSSTNIKKIQODKKG 476

RESULT 7  
US-09-815-242-13759  
Sequence 13759, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in Prokaryotes  
FILE REFERENCE: ELITRA.011A  
CURRENT FILING DATE: 2001-03-21  
PRIOR FILING DATE: 2000-03-21  
PRIOR FILING DATE: 2000-03-21  
PRIOR FILING DATE: 2000-05-23  
PRIOR FILING DATE: 2000-05-23  
PRIOR FILING DATE: 2000-05-26  
PRIOR FILING DATE: 2000-09-09  
PRIOR FILING DATE: 2000-10-23  
PRIOR FILING DATE: 2000-11-27  
PRIOR FILING DATE: 2000-12-22  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 13759  
LENGTH: 477  
TYPE: PRT  
ORGANISM: Salmonella typhi  
US-09-815-242-13759

Query Match 94.7%; Score 2265; DB 3; Length 477;  
Best Local Similarity 93.5%; Pred. No. 1.7e-187;  
Matches 445; Conservative 19; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MKVTLPEERAGVWVGVMDLDRYWGPTSRISPEAPVPVVKNTTIERPGGANVAMNI 60  
Db 1 MKVTLPEERAGVWVGVMDLDRYWGPTSRISPEAPVPVVKNTTIERPGGANVAMNI 60  
Qy 61 ASLGNARLVGLTGIDDAARALSKSLADVNVKDFVSVPTHTTTKRLVLSRNQOLRLD 120

```

Db 61 ASIGANARLVGLTGIDDAARLSKTLAEVNVKDFVSPHPHTTKARVSRNOQLRLD 120
Qy 121 FEESGFEVDPPQLPHERINQALSSIGALVLSDYAKGALASVQOMIOLARKAGVPLIDPKG 180
Db 121 FEESGFEVDPPQLPHERINQALSSIGALVLSDYAKGALASVQOMIOLARKAGVPLIDPKG 180
Qy 181 TDFERYGATLLTPNLSSEFAVAVGCKSEBELVERGKLIADYLSALVTRSEOGMTLL 240
Db 181 TDFERYGATLLTPNLSSEFAVAVGCKSEBELVERGKLIADYLSALVTRSEOGMTLL 240
Qy 241 QPKAPLHPMPTQAOEYVDVTGAGDTVIGVLAATLAAGNSLEACFPANAAAGVVGGLGT 300
Db 241 QPKAPLHPMPTQAOEYVDVTGAGDTVIGVLAATLAAGNSLEACFPANAAAGVVGGLGT 300
Qy 301 STVSPIELENAVRGRADTGFVMTSEBELKLAVAARRGKGVMTNGVPDILHAGVSYL 360
Db 301 STVSPIELENAVRGRADTGFVMTSEBELKLAVAARRGKGVMTNGVPDILHAGVSYL 360
Qy 361 ANARKLGDRLIVANSDDASTKRLKGSRPVNPLEQRMIVIGALESDVMVVSFEEDTPORL 420
Db 361 ANARKLGDRLIVANSDDASTKRLKGSRPVNPLEQRMIVIGALESDVMVVSFEEDTPORL 420
Qy 421 IAGILPDLVKGDDYKPEBIAGSKRVANGGEVLVNFEDCGSTTNIKKIQODK 476
Db 421 IAGILPDLVKGDDYKPEBIAGSKRVANGGEVLVNFEDCGSTTNIKKIQODES 476

```

```

RESULT 8
US-10-282-122A-75478
; Sequence 75478, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zykkind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282.122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 75478

```

```

; LENGTH: 477
; TYPE: PRT
; ORGANISM: Salmonella typhi
US-10-282-122A-75478
Query Match          94.7%; Score 2265; DB 4; Length 477;
Best Local Similarity 93.5%; Pred. No. 1,7e-187;
Matches 445; Conservative 19; Mismatches 12; Indels 0; Gaps 0;
Qy 1 MKYTLPEFERAGVNVGDVMDLRYWTGPTSRISPEAPVYVKNVNTIIEERGGANVAMNI 60
Db 1 MKYTLPEFERAGVNVGDVMDLRYWTGPTSRISPEAPVYVKNVNTIIEERGGANVAMNI 60
Qy 61 ASIGANARLVGLTGIDDAARLSKTLAEVNVKDFVSPHPHTTKARVSRNOQLRLD 120
Db 61 ASIGANARLVGLTGIDDAARLSKTLAEVNVKDFVSPHPHTTKARVSRNOQLRLD 120
Qy 121 FEESGFEVDPPQLPHERINQALSSIGALVLSDYAKGALASVQOMIOLARKAGVPLIDPKG 180
Db 121 FEESGFEVDPPQLPHERINQALSSIGALVLSDYAKGALASVQOMIOLARKAGVPLIDPKG 180
Qy 181 TDFERYGATLLTPNLSSEFAVAVGCKSEBELVERGKLIADYLSALVTRSEOGMTLL 240
Db 181 TDFERYGATLLTPNLSSEFAVAVGCKSEBELVERGKLIADYLSALVTRSEOGMTLL 240
Qy 241 QPKAPLHPMPTQAOEYVDVTGAGDTVIGVLAATLAAGNSLEACFPANAAAGVVGGLGT 300
Db 241 QPKAPLHPMPTQAOEYVDVTGAGDTVIGVLAATLAAGNSLEACFPANAAAGVVGGLGT 300
Qy 301 STVSPIELENAVRGRADTGFVMTSEBELKLAVAARRGKGVMTNGVPDILHAGVSYL 360
Db 301 STVSPIELENAVRGRADTGFVMTSEBELKLAVAARRGKGVMTNGVPDILHAGVSYL 360
Qy 361 ANARKLGDRLIVANSDDASTKRLKGSRPVNPLEQRMIVIGALESDVMVVSFEEDTPORL 420
Db 361 ANARKLGDRLIVANSDDASTKRLKGSRPVNPLEQRMIVIGALESDVMVVSFEEDTPORL 420
Qy 421 IAGILPDLVKGDDYKPEBIAGSKRVANGGEVLVNFEDCGSTTNIKKIQODK 476
Db 421 IAGILPDLVKGDDYKPEBIAGSKRVANGGEVLVNFEDCGSTTNIKKIQODES 476

```

```

RESULT 9
US-10-282-122A-60277
; Sequence 60277, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zykkind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282.122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09

```

```

; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent version 3.1
; SEQ ID NO: 60277
; LENGTH: 477
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-10-282-122A-60277

```

```

Query Match      93.1%; Score 2227; DB 4; Length 477;
Best Local Similarity 92.2%; Pred. No. 3.3e-184;
Matches 437; Conservative 25; Mismatches 12; Indels 0; Gaps 0;

```

```

Qy 1 MKVTLPFERAGVNVGDVMDLDRWYGPTRISPEAPVVKVNTTIERPGGANVAMNI 60
Db 1 MKVTLPFERAGVNVGDVMDLDRWYGPTRISPEAPVVKVNTTIERPGGANVAMNI 60
Qy 61 ASLGANALVGLTGIDDAARALSKSLADVNVKCDPVSPHTPTTKLRVLSRNOQLRLD 120
Db 61 ASLGANALVGLTGIDDAARALSKSLADVNVKCDPVSPHTPTTKLRVLSRNOQLRLD 120
Qy 121 FEEGFEVDPQLHERINQALSSIGALVLSDYAKGALASVQOMIQLARKAGVPLIDPKG 180
Db 121 FEEGFEVDPQLHERINQALSSIGALVLSDYAKGALASVQOMIQLARKAGVPLIDPKG 180
Qy 121 FEEGFEVDPQLHERINQALSSIGALVLSDYAKGALASVQOMIQLARKAGVPLIDPKG 180
Db 121 FEEGFEVDPQLHERINQALSSIGALVLSDYAKGALASVQOMIQLARKAGVPLIDPKG 180
Qy 181 TDFERYRATLTTPPLSFEEAVNGKCTEERIVERGKMLADYELSLALVTRSEQKSL 240
Db 181 TDFERYRATLTTPPLSFEEAVNGKCTEERIVERGKMLADYELSLALVTRSEQKSL 240
Qy 181 TDFERYRATLTTPPLSFEEAVNGKCTEERIVERGKMLADYELSLALVTRSEQKSL 240
Db 181 TDFERYRATLTTPPLSFEEAVNGKCTEERIVERGKMLADYELSLALVTRSEQKSL 240
Qy 241 QPGKAPLMPFOAQEVYVNTGAGPTVIGVLAATLAAGNSLEBACFPANAAAGVVGKLT 300
Db 241 QPGKAPLMPFOAQEVYVNTGAGPTVIGVLAATLAAGNSLEBACFPANAAAGVVGKLT 300
Qy 241 QPGKAPLMPFOAQEVYVNTGAGPTVIGVLAATLAAGNSLEBACFPANAAAGVVGKLT 300
Db 241 QPGKAPLMPFOAQEVYVNTGAGPTVIGVLAATLAAGNSLEBACFPANAAAGVVGKLT 300
Qy 301 STVPSEILENNAVGRADTFGVTMEELKLAVAAARKGKRVMTNGVFIIAHGHSYL 360
Db 301 STVPSEILENNAVGRADTFGVTMEELKLAVAAARKGKRVMTNGVFIIAHGHSYL 360
Qy 301 STVPSEILENNAVGRADTFGVTMEELKLAVAAARKGKRVMTNGVFIIAHGHSYL 360
Db 301 STVPSEILENNAVGRADTFGVTMEELKLAVAAARKGKRVMTNGVFIIAHGHSYL 360
Qy 361 ANARKLGRLLIVANVSDASTKRLKDSRPVNPLEQRMIVLGALBAVDVWVSFEEDTPQRL 420
Db 361 ANARKLGRLLIVANVSDASTKRLKDSRPVNPLEQRMIVLGALBAVDVWVSFEEDTPQRL 420
Qy 421 IAGILPDLVLKGGDYKPEEIASGEVWANGGEVLVLFNEDGCSSTNIKKIQ 474
Db 421 IAGILPDLVLKGGDYKPEEIASGEVWANGGEVLVLFNEDGCSSTNIKKIQ 474

```

```

RESULT 10
US-10-282-122A-73168
; Sequence 73168, Application US/10282122A
; Publication NO. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangou
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.

```

```

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent version 3.1
; SEQ ID NO: 73168
; LENGTH: 463
; TYPE: PRT
; ORGANISM: Salmonella paratyphi A
US-10-282-122A-73168

```

```

Query Match      91.8%; Score 2196; DB 4; Length 463;
Best Local Similarity 93.5%; Pred. No. 1.6e-181;
Matches 435; Conservative 18; Mismatches 10; Indels 2; Gaps 2;

```

```

Qy 8 FEEAGVNVGDVMDLDRWYGPTRISPEAPVVKVNTTIERPGGANVAMNIASIGANA 67
Db 1 FEEAGVNVGDVMDLDRWYGPTRISPEAPVVKVNTTIERPGGANVAMNIASIGANA 60
Qy 68 RLVGLGIDDAARALSKSLADVNVKCDPVSPHTPTTKLRVLSRNOQLRLDPEBEG 127
Db 61 RLVGLGIDDAARALSKSLADVNVKCDPVSPHTPTTKLRVLSRNOQLRLDPEBEG 119
Qy 128 VDPQLHERINQALSSIGALVLSDYAKGALASVQOMIQLARKAGVPLIDPKGTFERYR 187
Db 120 VDPQLHERINQALSSIGALVLSDYAKGALASVQOMIQLARKAGVPLIDPKGTFERYR 179
Qy 188 GATLTPNLSEFAVNVGKCTEERIVERGKMLADYELSLALVTRSEQKSLQPGKAPL 247
Db 180 GATLTPNLSEFAVNVGKCTEERIVERGKMLADYELSLALVTRSEQKSLQPGKAPL 239
Qy 248 HMPFOAQEVYVNTGAGPTVIGVLAATLAAGNSLEBACFPANAAAGVVGKLTSTVPIE 307
Db 240 HMPFOAQEVYVNTGAGPTVIGVLAATLAAGNSLEBACFPANAAAGVVGKLTSTVPIE 299
Qy 308 LENAVERGRADTFGVTMEELKLAVAAARKGKRVMTNGVFIIAHGHSYLANARKLG 367
Db 300 LENAVERGRADTFGVTMEELKLAVAAARKGKRVMTNGVFIIAHGHSYLANARKLG 358
Qy 368 DRLIVANVSDASTKRLKDSRPVNPLEQRMIVLGALBAVDVWVSFEEDTPQRLIAGILPD 427
Db 359 DRLIVANVSDASTKRLKDSRPVNPLEQRMIVLGALBAVDVWVSFEEDTPQRLIAGILPD 418
Qy 428 LTVKGGDYKPEEIASGEVWANGGEVLVLFNEDGCSSTNIKKIQ 472
Db 419 LTVKGGDYKPEEIASGEVWANGGEVLVLFNEDGCSSTNIKKIQ 463

```

```

RESULT 11
US-10-282-122A-78165
; Sequence 78165, Application US/10282122A

```



```
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zykkind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Foreyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 78165
LENGTH: 476
TYPE: PRT
ORGANISM: Yersinia pestis
US-10-282-122A-78165

Query Match      86.8%; Score 2076; DB 4; Length 476;
Best Local Similarity 86.1%; Pred. No. 4,2e-171;
Matches 409; Conservative 32; Mismatches 34; Indels 0; Gaps 0;

QY 1 MKVTLPEFRAGVNVGDVMDLDRYWGPTSRISPEAPVPVVKVNTIEERPGAAVAMNI 60
DB 1 MKVTLPEFRAGVNVGDVMDLDRYWGPTSRISPEAPVPVVKVNTIEERPGAAVAMNI 60
QY 61 ASIGANRLVGLTGIDDAARLSTSLADVNVKCDPFSVPVPHPTTKRLVSRNOQLRLD 120
DB 61 ASIGANRLVGLTGIDDAARLSTSLADVNVKCDPFSVPVPHPTTKRLVSRNOQLRLD 120
QY 121 FEEGFEQVDPOPLERINQALSSIGALVLSDYAGALASVOOMIQLARKAGVPYLIDPKG 180
DB 121 FEEGFEQVDPOPLERINQALSSIGALVLSDYAGALASVOOMIQLARKAGVPYLIDPKG 180
QY 181 TDFERYGATLLTNLSFEFAVVKCKTEEBIVERGKMLADYELSAALLVTRSEQMSLL 240
DB 181 TDFERYGATLLTNLSFEFAVVKCKTEEBIVERGKMLADYELSAALLVTRSEQMSLL 240
QY 241 QPGKAPLHMPQAOBVDVNGAGDPTVGVLAATLAAGNSLEAEPFANAAAGVVGKLGCT 300
DB 241 QPGKAPLHMPQAOBVDVNGAGDPTVGVLAATLAAGNSLEAEPFANAAAGVVGKLGCT 300
QY 301 STVSPLEENAVNRADDTGFGVMTSEEBIKLVAARKGKGVWNTNGVFDIILHAGHSYL 360
DB 301 STVSPLEENAVNRADDTGFGVMTSEEBIKLVAARKGKGVWNTNGVFDIILHAGHSYL 360
```

```
DB 301 STVSPLEENAVNRADDTGFGVMTSEEBIKLVAARKGKGVWNTNGVFDIILHAGHSYL 360
QY 361 ANARKGDRILVAVNSDASKRLKGSRPVNPLEOMITVIGALEAVDWMVSPEDTPORL 420
DB 361 ANARKGDRILVAVNSDASKRLKGSRPVNPLEOMITVIGALEAVDWMVSPEDTPORL 420
QY 421 IAGILPDLVKGADYKPEBIAGSKSEVWANGSEVLVNFEDGCTNIIKKIQODK 475
DB 421 IAGILPDLVKGADYKPEBIAGSKSEVWANGSEVLVNFEDGCTNIIKKIQODK 475

RESULT 12
US-10-282-122A-68720
Sequence 68720, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zykkind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Foreyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 68720
LENGTH: 474
TYPE: PRT
ORGANISM: Proteus mirabilis
US-10-282-122A-68720

Query Match      83.9%; Score 2007; DB 4; Length 474;
Best Local Similarity 82.2%; Pred. No. 4e-165;
Matches 389; Conservative 43; Mismatches 41; Indels 0; Gaps 0;

QY 1 MKVTLPEFRAGVNVGDVMDLDRYWGPTSRISPEAPVPVVKVNTIEERPGAAVAMNI 60
DB 1 MKVTLPEFRAGVNVGDVMDLDRYWGPTSRISPEAPVPVVKVNTIEERPGAAVAMNI 60
QY 61 ASIGANRLVGLTGIDDAARLSTSLADVNVKCDPFSVPVPHPTTKRLVSRNOQLRLD 120
DB 61 ASIGANRLVGLTGIDDAARLSTSLADVNVKCDPFSVPVPHPTTKRLVSRNOQLRLD 120
```

Qy 121 FEESPEGVDPOLHERINQALSTIGALVLSYDANGALASVOQMTQLARKAGVPLIDPKG 180  
Db 121 FEESFVSNSPEPIYERIQOALPSWGALILSDYAKGALSHVOQMTQLAKKGVPLIDPKG 180  
Qy 181 TDFERYGATLLTNTLSFEELVNGKCKTEERIVERGMKLIADYELSLVLTRESQMSIL 240  
Db 181 SDFERYKATLLTNTMSFEBOVGVGCHTDBELVERGTQVLDLADALLTRESRGSLL 240  
Qy 241 QPGKAPLHPTQAOEVDYDTGAGDTVIGVLAATLAAGNSLEBACFPANAAAGVVGKLG 300  
Db 241 QANAPLHLPTQAOEVDYDTGAGDTVIGVLAATLAAGNSLEBACFPANAAAGVVGKLG 300  
Qy 301 STVSPILLENAAVRGADTGFVMTBEEKLAVAARKEGKVMTNGVDTLHAGVSYL 360  
Db 301 STVSPILLENAAVRGADTGFVMTBEEKLAVAARKEGKVMTNGVDTLHAGVSYL 360  
Qy 361 ANARKLGRLLVAVNSDASTKRLKGDSPVNPLOMTVIGALAVDMVVSFEEDTQRL 420  
Db 361 ANARKLGRLLVAVNSDASTKRLKGDSPVNPLOMTVIGALAVDMVVSFEEDTQRL 420  
Qy 421 IAGILPDLVKGADYKPEELIAGSKVWANGGEVLVLPEDGCTTNIKKIQ 473  
Db 421 IASVLPDLVKGADYKPEELIAGSKVWANGGEVLVLPEDGCTTNIKKIK 473

RESULT 13  
US-10-282-122A-58533  
Sequence 58533, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl  
APPLICANT: Zykind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT FILING DATE: 2003-02-20  
CURRENT APPLICATION NUMBER: US/10/282,122A  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remainder Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: Patent version 3.1  
SEQ ID NO 58533  
LENGTH: 476  
TYPE: PRT  
ORGANISM: Haemophilus influenzae

US-10-282-122A-58533  
Query Match 71.5%; Score 1712; DB 4; Length 476;  
Best Local Similarity 71.0%; Pred. No. 1,6e-139;  
Matches 331; Conservative 58; Mismatches 77; Indels 0; Gaps 0;  
Qy 7 EFERAGVMTVDMDRYWGPTRISPEAPVNVKVTTEEPGGAANYANNIASIGAN 66  
Db 7 EFRQAKVLVLDGVMILRYWFGATNRISPAPVNVVQENEERAGGAANYANNIASINVP 66  
Qy 67 ARVGLTGIDDAARALSKSLADVNVKODFVSVTHPTITGLRVLSNQOILRLDFEEGE 126  
Db 67 VQMLGLIGDDETSALSLIBEKOKIDCNFVALTHTHTITKRLISRHQULRLDFEEDFN 126  
Qy 127 GVDPOPLHERINQALSTIGALVLSYDANGALASVOQMTQLARKAGVPLIDPKGTDFERY 186  
Db 127 NVDCOKLALLESAAVKNYGALILSDYKGTLLKOVQMTQLARKANVPVLPIDPKGTDFERY 186  
Qy 187 RGATLLTNPISERBAYVNGKCTEERIVERGMKLIADYELSLVLTRESQMSILQPKAP 246  
Db 187 RGATLLTNPMSERBAYVNGKCTEERIVERGMKLIADYELSLVLTRESQMSILRPQEP 246  
Qy 247 LHMPTQAOEVDYDTGAGDTVIGVLAATLAAGNSLEBACFPANAAAGVVGKLGSTVSP 306  
Db 247 YHLPVAKGVFDYDTGAGDTVIGVLAATLAAGNSFEESCYNANVAGVVGKLGSTVSTV 306  
Qy 307 ELENAVRGADTGFVMTBEEKLAVAARKEGKVMTNGVDTLHAGVSYLANARKL 366  
Db 307 ELENAIHAREPTFGIMSEBELDVAQAKAREKIVMTNGCGDILHPSHILLENARKL 366  
Qy 367 GDRLLVAVNSDASTKRLKGDSPVNPLOMTVIGALAVDMVVSFEEDTQRLIAGILP 426  
Db 367 GDRLLVAVNSDDSVKRLKGESRPINNLERNMAVLAGLASVDMVLPTEEDTQRLIGILP 426  
Qy 427 DILVKGADYKPEELIAGSKVWANGGEVLVLPEDGCTTNIKKIQ 472  
Db 427 DILVKGADYKPEELIAGSKVWANGGEVLVLPEDGCTTNIKKIK 472

RESULT 14  
US-10-282-122A-67124  
Sequence 67124, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl  
APPLICANT: Zykind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT FILING DATE: 2003-02-20  
CURRENT APPLICATION NUMBER: US/10/282,122A  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27  
 PRIOR APPLICATION NUMBER: 60/257,931  
 PRIOR FILING DATE: 2000-12-22  
 PRIOR APPLICATION NUMBER: 60/267,636  
 PRIOR FILING DATE: 2001-02-09  
 PRIOR APPLICATION NUMBER: 60/269,308  
 PRIOR FILING DATE: 2001-02-16  
 Remaining Prior Application data removed - See File Wrapper or PALM.  
 NUMBER OF SEQ ID NOS: 78614  
 SOFTWARE: PatentIn version 3.1  
 SEQ ID NO: 67124  
 LENGTH: 476  
 TYPE: PRT  
 ORGANISM: Pasteurella multocida  
 US-10-282-122A-67124

Query Match 71.1%; Score 1701; DB 4; Length 476;  
 Best Local Similarity 69.7%; Pred. No. 1.5e-138;  
 Matches 327; Conservative 65; Mismatches 77; Indels 0; Gaps 0;

QY 7 EFEEAGVNVVDVMDLDRYVGPTRISPEAPVPPVKNVTIEERGGANVAMNTASIGAN 66  
 DB 7 EFEEAGVNVVDVMDLDRYVGPTRISPEAPVPPVKNVTIEERGGANVAMNTASIGAN 66  
 QY 67 ARLVGLGIDDAABALSKSLADVNVKCPVSVPTHTTKLRVLSRNOQLRLDPEEGFE 126  
 DB 67 VKLGLTGRDERTGLAESLAKSRIBCDPVOJLHPTTKLRISRNOQLRLDPEEDFQ 126  
 QY 127 GVDDPQHERINQALSSIGALVLSDYAKGALASVOOMIQARAKGVPLIDPKGTDFERY 186  
 DB 127 NVTSDALQKLDALQNVGALVLSDYKGKTLNVOYKMIQARQANIPVLPIDPKGTDFERY 186  
 QY 187 RGATLTPNLSFEFAVVGKCTEERIEVERGKLLADYELSLVTRSEOGSLIQPGAP 246  
 DB 187 RGATLTPNLSFEFAVVGKCTEERIEVERGKLLADYELSLVTRSEOGSLIQPGAP 246  
 QY 247 LHMTOAOEYVDVAGDPTVIGVLAATLAAGNSLEBACFPANAAAGVVGGLGTSTSP 306  
 DB 247 PHLTKEKVDVAGDPTVIGVLAATLAAGNSLEBACFPANAAAGVVGGLGTSTSP 306  
 QY 307 ELENAVGRADTGFVMTBEELKLAVAARRGKRVMTNGVFDILHAGHVSYLENARKL 366  
 DB 307 ELENAVGRADTGFVMTBEELKLAVAARRGKRVMTNGVFDILHAGHVSYLENARKL 366  
 QY 367 GDRILIVANSDASRKLGSRPNVPLEQRMTIVGALBAVDVVSFEEDTPQRLIAGILP 426  
 DB 367 GDRILIVANSDASRKLGSRPNVPLEQRMTIVGALBAVDVVSFEEDTPQRLIAGILP 426  
 QY 427 DLVKGDDYKPEEIASGKEVWANGGEVVLNFEEDGCTTNIKKIQODK 475  
 DB 427 DLVKGDDYKPEEIASGKEVWANGGEVVLNFEEDGCTTNIKKIQODK 475

RESULT 15  
 US-10-282-122A-69312  
 Sequence 69312, Application US/102821222A  
 Publication No. US20040029129A1  
 GENERAL INFORMATION:  
 APPLICANT: Wang, Liangsu  
 APPLICANT: Zamudio, Carlos  
 APPLICANT: Malone, Cheryl  
 APPLICANT: Haselbeck, Robert  
 APPLICANT: Ohlsen, Karl  
 APPLICANT: Zykend, Judith  
 APPLICANT: Wall, Daniel  
 APPLICANT: Trawick, John  
 APPLICANT: Carr, Grant  
 APPLICANT: Yamamoto, Robert  
 APPLICANT: Forsyth, R.  
 APPLICANT: Xu, H.  
 TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
 FILE REFERENCE: EUTRA.034A  
 CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20  
 PRIOR APPLICATION NUMBER: 60/191,078  
 PRIOR FILING DATE: 2000-03-21  
 PRIOR APPLICATION NUMBER: 60/206,848  
 PRIOR FILING DATE: 2000-05-23  
 PRIOR APPLICATION NUMBER: 60/207,727  
 PRIOR FILING DATE: 2000-05-26  
 PRIOR APPLICATION NUMBER: 60/220,335  
 PRIOR FILING DATE: 2000-09-06  
 PRIOR APPLICATION NUMBER: 60/230,347  
 PRIOR FILING DATE: 2000-09-09  
 PRIOR APPLICATION NUMBER: 60/242,578  
 PRIOR FILING DATE: 2000-10-23  
 PRIOR APPLICATION NUMBER: 60/253,625  
 PRIOR FILING DATE: 2000-11-27  
 PRIOR APPLICATION NUMBER: 60/257,931  
 PRIOR FILING DATE: 2000-12-22  
 PRIOR APPLICATION NUMBER: 60/267,636  
 PRIOR FILING DATE: 2001-02-09  
 PRIOR APPLICATION NUMBER: 60/269,308  
 Remaining Prior Application data removed - See File Wrapper or PALM.  
 NUMBER OF SEQ ID NOS: 78614  
 SOFTWARE: PatentIn version 3.1  
 SEQ ID NO: 69312  
 LENGTH: 474  
 TYPE: PRT  
 ORGANISM: Pseudomonas syringae  
 US-10-282-122A-69312

Query Match 56.7%; Score 1357.5; DB 4; Length 474;  
 Best Local Similarity 58.1%; Pred. No. 9.3e-109;  
 Matches 274; Conservative 70; Mismatches 127; Indels 1; Gaps 1;

QY 1 MKYTLPEFEAGVNVVDVMDLDRYVGPTRISPEAPVPPVKNVTIEERGGANVAMNT 60  
 DB 1 MKYTLPEFEAGVNVVDVMDLDRYVGPTRISPEAPVPPVKNVTIEERGGANVAMNT 60  
 QY 61 ASLGAARLVGLGIDDAABALSKSLADVNVKCPVSVPTHTTKLRVLSRNOQLRLD 120  
 DB 61 ASLGAARLVGLGIDDAABALSKSLADVNVKCPVSVPTHTTKLRVLSRNOQLRLD 120  
 QY 121 FESEFEGVDPPHERINQALSSIGALVLSDYAKGALASVOOMIQARAKGVPLIDPKG 180  
 DB 121 FESEFEGVDPPHERINQALSSIGALVLSDYAKGALASVOOMIQARAKGVPLIDPKG 180  
 QY 181 TDPERYRGATLTPNLSFEFAVVGKCTEERIEVERGKLLADYELSLVTRSEOGSL 240  
 DB 181 TDPERYRGATLTPNLSFEFAVVGKCTEERIEVERGKLLADYELSLVTRSEOGSL 240  
 QY 241 QPGKAPLHMTOAOEYVDVAGDPTVIGVLAATLAAGNSLEBACFPANAAAGVVGGLGT 300  
 DB 241 QPGKAPLHMTOAOEYVDVAGDPTVIGVLAATLAAGNSLEBACFPANAAAGVVGGLGT 300  
 QY 301 STVSPTELENAVGRADTGFVMTBEELKLAVAARRGKRVMTNGVFDILHAGHVSYLE 360  
 DB 301 STVSPTELENAVGRADTGFVMTBEELKLAVAARRGKRVMTNGVFDILHAGHVSYLE 360  
 QY 361 ANARKLGDRLIVANSDASRKLGSRPNVPLEQRMTIVGALBAVDVVSFEEDTPQRL 420  
 DB 361 ANARKLGDRLIVANSDASRKLGSRPNVPLEQRMTIVGALBAVDVVSFEEDTPQRL 420  
 QY 421 IAGILPDLVKGDDYKPEEIASGKEVWANGGEVVLNFEEDGCTTNIKKIQ 472  
 DB 421 IAGILPDLVKGDDYKPEEIASGKEVWANGGEVVLNFEEDGCTTNIKKIQ 472

Search completed: March 18, 2006, 06:21:19  
 Job time: 170 secs

**THIS PAGE LEFT BLANK**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model1

Run on: March 19, 2006, 11:24:39 ; Search time 2093 Seconds

(without alignments)  
5665.689 Million cell updates/sec

Title: US-10-520-820-28

Perfect score: 1434  
Sequence: 1 atgaaagtaacgtcgcaga.....aacagataaaagctaa 1434

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA Main:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US11\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1434	100.0	1434	US-09-741-669-276	Sequence 276, App
2	1434	100.0	1434	US-09-912-020-165	Sequence 165, App
3	1434	100.0	1434	US-09-815-242-6204	Sequence 6204, Ap
4	1434	100.0	1434	US-10-282-122A-6430	Sequence 6430, Ap
5	1434	100.0	1434	US-10-771-241-165	Sequence 165, App
6	1018.2	71.0	1428	US-10-282-122A-20018	Sequence 20018, A
7	1008.6	70.3	1434	US-09-815-242-6651	Sequence 9651, Ap
8	1008.6	70.3	1434	US-10-282-122A-39294	Sequence 39294, A
9	1007	70.2	1431	US-10-282-122A-24093	Sequence 24093, A
10	935.2	65.1	1397	US-10-282-122A-36984	Sequence 36984, A
11	879.2	61.3	1431	US-10-282-122A-41981	Sequence 41981, A
12	725	50.6	1422	US-10-282-122A-32536	Sequence 32536, A
13	635.2	44.3	1431	US-10-282-122A-22349	Sequence 22349, A
14	635.2	44.3	1830121	US-10-329-670-1	Sequence 1, Appl
15	635.2	44.3	1830121	US-10-158-865-1	Sequence 1, Appl
16	635.2	44.3	1830121	US-10-981-687-1	Sequence 1, Appl
17	625.6	43.6	908766	US-10-795-159-685	Sequence 685, App
18	617.4	43.1	1431	US-10-282-122A-30940	Sequence 30940, A
19	507.8	35.4	707	US-10-363-345A-7681	Sequence 7681, Ap
20	507.8	35.4	707	US-10-363-345A-7682	Sequence 7682, Ap
21	507.8	35.4	707	US-10-363-345A-7682	Sequence 7682, Ap
22	507.8	35.4	707	US-10-363-345A-7682	Sequence 7682, Ap
23	497.8	34.7	1425	US-09-815-242-7981	Sequence 7981, Ap

24	497.8	34.7	1425	US-10-282-122A-30585	Sequence 30585, A
25	493	34.4	1422	US-10-282-122A-33128	Sequence 33128, A
26	465.6	32.5	1407	US-10-282-122A-31665	Sequence 31665, A
27	425.6	29.7	707	US-10-363-345A-7683	Sequence 7683, Ap
28	425.2	29.7	707	US-10-363-345A-7684	Sequence 7684, Ap
29	425.2	29.7	707	US-10-363-345A-7684	Sequence 7684, Ap
30	425.2	29.7	707	US-10-363-345A-7684	Sequence 7684, Ap
31	406.2	28.3	2055	US-10-450-763-25673	Sequence 25673, A
32	397	27.7	555	US-10-363-345A-32035	Sequence 32035, A
33	397	27.7	555	US-10-363-345A-32036	Sequence 32036, A
34	397	27.7	555	US-10-363-345A-32036	Sequence 32036, A
35	397	27.7	555	US-10-363-345A-32036	Sequence 32036, A
36	390.4	27.2	1001	US-10-343-561-64	Sequence 64, Appl
37	388.6	27.1	3641	US-10-795-159-458	Sequence 458, App
38	340.8	23.8	555	US-10-363-345A-32033	Sequence 32033, A
39	340.8	23.8	555	US-10-363-345A-32034	Sequence 32034, A
40	340.8	23.8	555	US-10-363-345A-32033	Sequence 32033, A
41	340.8	23.8	555	US-10-363-345A-32034	Sequence 32034, A
42	302.6	21.1	706	US-10-363-345A-7297	Sequence 7297, Ap
43	302.6	21.1	706	US-10-363-345A-7298	Sequence 7298, Ap
44	302.6	21.1	706	US-10-363-345A-7297	Sequence 7297, Ap
45	302.6	21.1	706	US-10-363-345A-7298	Sequence 7298, Ap

## ALIGNMENTS

RESULT 1  
US-09-741-669-276  
Sequence 276, Application US/09741669  
Patent No. US20020022718A1  
GENERAL INFORMATION:  
APPLICANT: Forsyth, R. Allyn  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zyskind, Judith W.  
TITLE OF INVENTION: Genes identified as required for  
TITLE OF INVENTION: Proliferation of E. coli  
FILE REFERENCE: ELITRA 009A  
CURRENT APPLICATION NUMBER: US/09/741.669  
CURRENT FILING DATE: 2000-12-19  
PRIOR APPLICATION NUMBER: US 60/173005  
NUMBER OF SEQ ID NOS: 481  
SOFTWARE: PaastSeq for Windows Version 4.0  
SEQ ID NO 276  
LENGTH: 1434  
TYPE: DNA  
ORGANISM: Escherichia coli  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)...(1434)  
US-09-741-669-276  
Query Match 100.0%; Score 1434; DB 3; Length 1434;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGAAAGTAACGCTGCCAGAGTTGAACGTGCAGAGTATGCTGTTGATGATG 60  
1 ATGAAAGTAACGCTGCCAGAGTTGAACGTGCAGAGTATGCTGTTGATGATG 60  
Db 1 ATGAAAGTAACGCTGCCAGAGTTGAACGTGCAGAGTATGCTGTTGATGATG 60  
QY 61 CTGAGTCCTGATCGTATCGGAGCCGACAGTCGATCTCGCGAAGCGCGGCGCG 120  
61 CTGAGTCCTGATCGTATCGGAGCCGACAGTCGATCTCGCGAAGCGCGGCGCG 120  
Db 61 CTGAGTCCTGATCGTATCGGAGCCGACAGTCGATCTCGCGAAGCGCGGCGCG 120  
QY 121 GTTAAAGTAAATACATCGAAGACGTCCGGCGCGGCTTAACGTGCGATGAATATC 180  
121 GTTAAAGTAAATACATCGAAGACGTCCGGCGCGGCTTAACGTGCGATGAATATC 180  
Db 121 GTTAAAGTAAATACATCGAAGACGTCCGGCGCGGCTTAACGTGCGATGAATATC 180  
QY 181 GCTTCTCTCGTCTATATGACCGCTGCTGCTGTTGACGGGCAATTGACATGACGCGCGC 240  
181 GCTTCTCTCGTCTATATGACCGCTGCTGCTGTTGACGGGCAATTGACATGACGCGCGC 240  
Db 181 GCTTCTCTCGTCTATATGACCGCTGCTGCTGTTGACGGGCAATTGACATGACGCGCGC 240

QY 241 GCGGTAGTAAATCTCTGCGCGACGTCAACGTCAAAATGCGACTTCTGTTCTGTAACGACG 300  
DB 241 GCGGTAGTAAATCTCTGCGCGACGTCAACGTCAAAATGCGACTTCTGTTCTGTAACGACG 300  
QY 301 CATCCGACCATTTACCAATTAACGGGTACTTTCCGCGAACCAACAGCTGATCCGTCTGAT 360  
DB 301 CATCCGACCATTTACCAATTAACGGGTACTTTCCGCGAACCAACAGCTGATCCGTCTGAT 360  
QY 361 TTGGAAGAAGGTTTGAAGGTTGATCCGACGCGGTGACGAGCGGATTTAATCAGGCG 420  
DB 361 TTGGAAGAAGGTTTGAAGGTTGATCCGACGCGGTGACGAGCGGATTTAATCAGGCG 420  
QY 421 CTGAGTTGATTTGCGCGCTGCTGCTTCTGTAACGCAAGGTGCGCTGCGCAAGCGTA 480  
DB 421 CTGAGTTGATTTGCGCGCTGCTGCTTCTGTAACGCAAGGTGCGCTGCGCAAGCGTA 480  
QY 481 CAGCAGATGATCCACTGCGCGGTAAAGCGGGTTCGGTGTGATTTGATCCAAAGGT 540  
DB 481 CAGCAGATGATCCACTGCGCGGTAAAGCGGGTTCGGTGTGATTTGATCCAAAGGT 540  
QY 541 ACCGATTTTGAAGGCTACCGCGCGGTACGCTGTTAAGCGCGATCTCTGCGAATTTGAA 600  
DB 541 ACCGATTTTGAAGGCTACCGCGCGGTACGCTGTTAAGCGCGATCTCTGCGAATTTGAA 600  
QY 601 GCTGTTCTGCTAAATGTAAGACCGAAGAGATTTGTAAGCGCGCATGAACTGAT 660  
DB 601 GCTGTTCTGCTAAATGTAAGACCGAAGAGATTTGTAAGCGCGCATGAACTGAT 660  
QY 661 GCGCATTAACGAATCTCTGCTCTGTTAGTGAACCGCTTCCGAACGGGTATCCCTGCTG 720  
DB 661 GCGCATTAACGAATCTCTGCTCTGTTAGTGAACCGCTTCCGAACGGGTATCCCTGCTG 720  
QY 721 CAACCGGGTAAAGGCGCGCTGATATGCAACCGAAGCGGAGAGTGAAGTGAAGTGAAC 780  
DB 721 CAACCGGGTAAAGGCGCGCTGATATGCAACCGAAGCGGAGAGTGAAGTGAAGTGAAC 780  
QY 781 GGTGCGGCGCAACAGGTGATTTGCGCTCTGCGCGCAACGCTGCGAGCGGGTAAATTCGCTG 840  
DB 781 GGTGCGGCGCAACAGGTGATTTGCGCTCTGCGCGCAACGCTGCGAGCGGGTAAATTCGCTG 840  
QY 841 GAAGAAGCTCTCTCTTTGCGCAATGCGCGCGCTGCGGTGTGTGCGCAAACTGGGAAC 900  
DB 841 GAAGAAGCTCTCTCTTTGCGCAATGCGCGCGCTGCGGTGTGTGCGCAAACTGGGAAC 900  
QY 901 TCACGAGTTTGCAGGATCGAGCTGGAATAATCTGTAACGTGACGTCGATCAAGCTTT 960  
DB 901 TCACGAGTTTGCAGGATCGAGCTGGAATAATCTGTAACGTGACGTCGATCAAGCTTT 960  
QY 961 GCGGTGATGACCGAAGAGAACTGAAGCTGCGGTAGCGGACGCGGTAAACGTGTGAA 1020  
DB 961 GCGGTGATGACCGAAGAGAACTGAAGCTGCGGTAGCGGACGCGGTAAACGTGTGAA 1020  
QY 1021 AAAAGTGTGATGACCAAGGTGTCTTTGACATCTCTGACGCGCGGACAGTCTTTATCTG 1080  
DB 1021 AAAAGTGTGATGACCAAGGTGTCTTTGACATCTCTGACGCGCGGACAGTCTTTATCTG 1080  
QY 1081 GCAAAATGCCCGAAGCTGGGTGACGCTTTGATTTGTCGCTCAACAGGATGCTTCAAC 1140  
DB 1081 GCAAAATGCCCGAAGCTGGGTGACGCTTTGATTTGTCGCTCAACAGGATGCTTCAAC 1140  
QY 1141 AAACGCGTGAAGAGGGATTTCCGCGCGGTAAACCACTGAAACGCGTATGATGTGCTG 1200  
DB 1141 AAACGCGTGAAGAGGGATTTCCGCGCGGTAAACCACTGAAACGCGTATGATGTGCTG 1200  
QY 1201 GGGCGCATGGAAGCGGTGACCTGGGTATGTCGTTGAAGAGACAGCGCGCACGCTTG 1260  
DB 1201 GGGCGCATGGAAGCGGTGACCTGGGTATGTCGTTGAAGAGACAGCGCGCACGCTTG 1260  
QY 1261 ATTCGCGGAGTCTTGCAGATCTGTGTTGAAGAGCGCGCATTTAAACGAGAAAGATT 1320  
DB 1261 ATTCGCGGAGTCTTGCAGATCTGTGTTGAAGAGCGCGCATTTAAACGAGAAAGATT 1320  
QY 1321 GCCGGAAGTAAAGAGTCTGCGGCAACGCTGCGGCAAGTGTGTGCTCAACTTTGAAAGC 1380

DB 1321 GCGGGAAGTAAAGAGTCTGCGGCAACGCTGCGGCAAGTGTGTGCTCAACTTTGAAAGC 1380  
QY 1381 GGTGCTCGACGACCAACATCATCAAGATCAACAGGATTAAGGCTAA 1434  
DB 1381 GGTGCTCGACGACCAACATCATCAAGATCAACAGGATTAAGGCTAA 1434

RESULT 2  
US-09-912-020-165  
Sequence 165, Application US/09912020  
Patent No. US2002004592A1  
GENERAL INFORMATION:  
APPLICANT: Zykend, Judith  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Trawick, John  
APPLICANT: Forsyth, R. Allyn  
APPLICANT: Froelich, Jamie M.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN  
TITLE OF INVENTION: ESCHERICHIA COLI  
FILE REFERENCE: EUTRA.001DV1  
CURRENT APPLICATION NUMBER: US/09/912,020  
CURRENT FILING DATE: 2001-07-23  
PRIORITY APPLICATION NUMBER: 09/492,709  
PRIORITY FILING DATE: 2000-01-27  
PRIORITY APPLICATION NUMBER: 60/117,405  
PRIORITY FILING DATE: 1999-01-27  
NUMBER OF SEQ ID NOS: 485  
SOFTWARE: FASTSEQ for Windows Version 3.0  
SEQ ID NO 165  
LENGTH: 1434  
TYPE: DNA  
ORGANISM: E. Coli  
US-09-912-020-165

Query Match 100.0%; Score 1434; DB 3; Length 1434;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAAGTAAACGCTGCCAGATTGAACGTGACGAGTGAATGTTGTTGATGATGATG 60  
DB 1 ATGAAGTAAACGCTGCCAGATTGAACGTGACGAGTGAATGTTGTTGATGATGATG 60  
QY 61 CTGATGCTTAACGTGATCGGCGCCCAACAGTGTATCTGCGCGAAGCGCGGTGCCG 120  
DB 61 CTGATGCTTAACGTGATCGGCGCCCAACAGTGTATCTGCGCGAAGCGCGGTGCCG 120  
QY 121 GTTAAAGTAAATCAATCGAAGACGTCCGCGCGCGCGCTTAACGTGCGATGATATC 180  
DB 121 GTTAAAGTAAATCAATCGAAGACGTCCGCGCGCGCGCTTAACGTGCGATGATATC 180  
QY 181 GCTTCTCTGCTGCTAATGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240  
DB 181 GCTTCTCTGCTGCTAATGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240  
QY 241 GCGGTGATTAATCTCTGCGCGAAGTCAACGTGAATGCACTTCTGTTTGAACGAG 300  
DB 241 GCGGTGATTAATCTCTGCGCGAAGTCAACGTGAATGCACTTCTGTTTGAACGAG 300  
QY 301 CATCCGACCATTTACCAATTAACGGGTACTTTCCGCGAACCAACAGCTGATCCGTCTGAT 360  
DB 301 CATCCGACCATTTACCAATTAACGGGTACTTTCCGCGAACCAACAGCTGATCCGTCTGAT 360  
QY 361 TTGGAAGAAGGTTTGAAGGTTGATCCGACGCGGTGACGAGCGGATTTAATCAGGCG 420  
DB 361 TTGGAAGAAGGTTTGAAGGTTGATCCGACGCGGTGACGAGCGGATTTAATCAGGCG 420  
QY 421 CTGAGTTGATTTGCGCGCTGCTGCTTCTGTAACGCAAGGTGCGCTGCGCAAGCGTA 480  
DB 421 CTGAGTTGATTTGCGCGCTGCTGCTTCTGTAACGCAAGGTGCGCTGCGCAAGCGTA 480

481 CAGCAGATGATCCAACTGGCGGCTAAAGCGGGTGTCCGGGCTGATTTGATCCAAAGT 540  
 481 CAGCAGATGATCCAACTGGCGGCTAAAGCGGGTGTCCGGGCTGATTTGATCCAAAGT 540  
 541 ACCGATTTTGAAGCGCTACCGCGGCGCTACGCTGTTTAAAGCGGATCTCTCGAATTTGAA 600  
 541 ACCGATTTTGAAGCGCTACCGCGGCGCTACGCTGTTTAAAGCGGATCTCTCGAATTTGAA 600  
 601 GCTGTGTGCTGATTAATGATGACCGAAGAGATTTGTTGAGCGCGGCTGATTAATGATTA 660  
 601 GCTGTGTGCTGATTAATGATGACCGAAGAGATTTGTTGAGCGCGGCTGATTAATGATTA 660  
 661 GCGGATTAACGAACCTGGGCTCTGTGATGACCGGCTTCCGAAACAGGGATGTCGCTG 720  
 661 GCGGATTAACGAACCTGGGCTCTGTGATGACCGGCTTCCGAAACAGGGATGTCGCTG 720  
 721 CAAACCGGCTAAAGCGCGCTGCTGATTAAGCAACCAAGCGGAGAGATGATGACGTTACC 780  
 721 CAAACCGGCTAAAGCGCGCTGCTGATTAAGCAACCAAGCGGAGAGATGATGACGTTACC 780  
 781 GGTGCGGCGGCAACGCTGATTTGGGCTCTGCGGCAACGCTGCGAGCGGCTAATTCGCTG 840  
 781 GGTGCGGCGGCAACGCTGATTTGGGCTCTGCGGCAACGCTGCGAGCGGCTAATTCGCTG 840  
 841 GAAGAAGCTGCTCTTTTCCAAATGCGGCGGCTGCGGCTGCTGCTGCTGCTGCTGCTG 900  
 841 GAAGAAGCTGCTCTTTTCCAAATGCGGCGGCTGCGGCTGCTGCTGCTGCTGCTGCTG 900  
 901 TCCACGCTTTCCGCGCTGCTGATTAAGCAACCAAGCGGAGAGATGATGACGTTACC 960  
 901 TCCACGCTTTCCGCGCTGCTGATTAAGCAACCAAGCGGAGAGATGATGACGTTACC 960  
 961 GCGCTGATGACGAAAGAGAACTGAAAGTGGCGCTGAGCGGCGGCTAATTCGCTG 1020  
 961 GCGCTGATGACGAAAGAGAACTGAAAGTGGCGCTGAGCGGCGGCTAATTCGCTG 1020  
 1021 AAGTGTGATGACCAACGCTGCTTTTGAACATCTGCAACCGCGGCGGCTGCTGCTGCTG 1080  
 1021 AAGTGTGATGACCAACGCTGCTTTTGAACATCTGCAACCGCGGCGGCTGCTGCTGCTG 1080  
 1081 GCAAAATGCGCGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140  
 1081 GCAAAATGCGCGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140  
 1141 AAGCGCTGAAAGGGAATTCGCGCGGCTAATTCGCAACGCTGCAACGCTGATGATGCTG 1200  
 1141 AAGCGCTGAAAGGGAATTCGCGCGGCTAATTCGCAACGCTGCAACGCTGATGATGCTG 1200  
 1201 GCGGCACTGAAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260  
 1201 GCGGCACTGAAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260  
 1261 ATGCGCGGAGATCTTGGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320  
 1261 ATGCGCGGAGATCTTGGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320  
 1321 GCGCGGAGATTAAGAGTCTGCGGCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380  
 1321 GCGCGGAGATTAAGAGTCTGCGGCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380  
 1381 GCTGTGCTGCAACCAATCATCAAGAGATTCGAACGATTAAGAGCTTA 1434  
 1381 GCTGTGCTGCAACCAATCATCAAGAGATTCGAACGATTAAGAGCTTA 1434

RESULT 3  
 US-09-815-242-6204  
 Sequence 6204, Application US/09815242  
 Patent No. US20020061569A1  
 GENERAL INFORMATION:  
 APPLICANT: Haseelbeck, Robert  
 APPLICANT: Ohlsen, Karl L.

APPLICANT: Zykand, Judith W.  
 APPLICANT: Wall, Daniel  
 APPLICANT: Trawick, John D.  
 APPLICANT: Carr, Grant J.  
 APPLICANT: Yamamoto, Robert T.  
 APPLICANT: Xu, H. Howard  
 TITLE OF INVENTION: Identification of Essential Genes in  
 TITLE OF INVENTION: Prokaryotes  
 FILE REFERENCE: ELITRA-011A  
 CURRENT APPLICATION NUMBER: US/09/815,242  
 PRIOR FILING DATE: 2001-03-21  
 PRIOR APPLICATION NUMBER: 60/191,078  
 PRIOR FILING DATE: 2000-03-21  
 PRIOR APPLICATION NUMBER: 60/206,848  
 PRIOR FILING DATE: 2000-05-23  
 PRIOR APPLICATION NUMBER: 60/207,727  
 PRIOR FILING DATE: 2000-05-26  
 PRIOR APPLICATION NUMBER: 60/242,578  
 PRIOR FILING DATE: 2000-10-23  
 PRIOR APPLICATION NUMBER: 60/253,625  
 PRIOR FILING DATE: 2000-11-27  
 PRIOR APPLICATION NUMBER: 60/257,931  
 PRIOR FILING DATE: 2000-12-22  
 PRIOR APPLICATION NUMBER: 60/269,308  
 PRIOR FILING DATE: 2001-02-16  
 NUMBER OF SEQ ID NOS: 1410  
 SOFTWARE: FASTSEQ for Windows Version 4.0  
 SEQ ID NO 6204  
 LENGTH: 1434  
 TYPE: DNA  
 ORGANISM: Escherichia coli  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (1)...(1434)  
 US-09-815-242-6204

Query Match 100.0%; Score 1434; DB 3; Length 1434;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGAAAGTAAACGCTGCGCAGAGTTTGAACGTCAGAGATGATGCTGATGATGATG 60  
 1 ATGAAAGTAAACGCTGCGCAGAGTTTGAACGTCAGAGATGATGCTGATGATGATG 60  
 61 CTGATGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120  
 61 CTGATGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120  
 121 GTTAAAGTAAACGCTGCGCAGAGTTTGAACGTCAGAGATGATGCTGATGATGATG 180  
 121 GTTAAAGTAAACGCTGCGCAGAGTTTGAACGTCAGAGATGATGCTGATGATGATG 180  
 181 GCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240  
 181 GCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240  
 241 GCGCTGATTAATCTCTGCGGCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300  
 241 GCGCTGATTAATCTCTGCGGCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300  
 241 GCGCTGATTAATCTCTGCGGCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300  
 241 GCGCTGATTAATCTCTGCGGCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300  
 301 CATCCGACATTAACCAATTAAGGAGTCTTCCGCAACCAACGATGATCCGCTGAT 360  
 301 CATCCGACATTAACCAATTAAGGAGTCTTCCGCAACCAACGATGATCCGCTGAT 360  
 361 TTGAAAGAGTCTGCAAGGTTGATTCGCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420  
 361 TTGAAAGAGTCTGCAAGGTTGATTCGCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420  
 421 CTGATGCTGATTTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480  
 421 CTGATGCTGATTTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480  
 481 CAGCAGATGATCCAACTGGCGGCTAAAGCGGGTGTCCGGGCTGATTTGATCCAAAGT 540

Db	481	CAGCAGATGATCTCAACTGCGCCGCTGTAAGCCGGGTGTTCCGGTGTCTGATTGATCTCAAAAGCT	540
Qy	541	ACCGATTTTGAAGCGCTACCGCGCGGCTACGCTGTTAAACGCCGAATCTCTCGAAATTTGAA	600
Db	541	ACCGATTTTGAAGCGCTACCGCGCGGCTACGCTGTTAAACGCCGAATCTCTCGAAATTTGAA	600
Qy	601	GCTGTGTGCGGTTAAATGTAAACCGAACAAGAGATTTGTGAAGCGCGGCAATGAACTGATT	660
Db	601	GCTGTGTGCGGTTAAATGTAAACCGAACAAGAGATTTGTGAAGCGCGGCAATGAACTGATT	660
Qy	661	GCCGATTACGAACCTCTCGGCTCTGTATTGACCCGGTTCGAAACAGGGATGTGCTGCTG	720
Db	661	GCCGATTACGAACCTCTCGGCTCTGTATTGACCCGGTTCGAAACAGGGATGTGCTGCTG	720
Qy	721	CAACCGGGGTAAAGCGCCGCTGCAATGCGCAACCCAGGCGCAGGAAGTGTATGAGCTTACC	780
Db	721	CAACCGGGGTAAAGCGCCGCTGCAATGCGCAACCCAGGCGCAGGAAGTGTATGAGCTTACC	780
Qy	781	GATGCGGGCGACACGCGTATTTGCGGCTCTGGGCGGCAACGCTGCGACGCGGATTAATTCGCTG	840
Db	781	GATGCGGGCGACACGCGTATTTGCGGCTCTGGGCGGCAACGCTGCGACGCGGATTAATTCGCTG	840
Qy	841	GAAGAAGCCTGCTCTTCTTTCGCCAATGCGCGCGCTGCGTGGTGTGCGAATCTGGAAC	900
Db	841	GAAGAAGCCTGCTCTTCTTTCGCCAATGCGCGCGCTGCGTGGTGTGCGAATCTGGAAC	900
Qy	901	TCCACGGTTTTGGCCGATGCGAGCTGGAAAAATGCTGTACGTGACCGTGCAGATACAGGCTTT	960
Db	901	TCCACGGTTTTGGCCGATGCGAGCTGGAAAAATGCTGTACGTGACCGTGCAGATACAGGCTTT	960
Qy	961	GAGCGATGACCGGAAGAGAACTGAAAGCTGGCCGTAGCGGCGACGCGTAAACGTGTGAA	1020
Db	961	GAGCGATGACCGGAAGAGAACTGAAAGCTGGCCGTAGCGGCGACGCGTAAACGTGTGAA	1020
Qy	1021	AAAGTGTGATGACCAACGCTGTCTTTTGAACATCTGCAACGCCCGGCAACGTCTTATCTG	1080
Db	1021	AAAGTGTGATGACCAACGCTGTCTTTTGAACATCTGCAACGCCCGGCAACGTCTTATCTG	1080
Qy	1081	GCAATATGCCCGCAAGCTGGGTGACCGCGTGAATGTTGTGCGTCAACAGGAGATGCCCTCAC	1140
Db	1081	GCAATATGCCCGCAAGCTGGGTGACCGCGTGAATGTTGTGCGTCAACAGGAGATGCCCTCAC	1140
Qy	1141	AAACGGCTGAAAGGGGATTTCCCGCCCGGTAAACCCACTCGAACAGCGTATGATTTGCTG	1200
Db	1141	AAACGGCTGAAAGGGGATTTCCCGCCCGGTAAACCCACTCGAACAGCGTATGATTTGCTG	1200
Qy	1201	GCGCGACTCGGAAGCCGCTGACTGGTATGTGTCTTTGAAGAGACACCCCGCAGGCTTG	1260
Db	1201	GCGCGACTCGGAAGCCGCTGACTGGTATGTGTCTTTGAAGAGACACCCCGCAGGCTTG	1260
Qy	1261	ATGCGCGGGAATCTTGTCCAGATCTGTGTGTGTAAGGCGCGCACTATAAACAGAAAGATT	1320
Db	1261	ATGCGCGGGAATCTTGTCCAGATCTGTGTGTGTAAGGCGCGCACTATAAACAGAAAGATT	1320
Qy	1321	GCCGGAGTAAAGAGTCTGCGGCCAACGCTGCGGAAGTGTGTGCTCAACTTTGAAGAC	1380
Db	1321	GCCGGAGTAAAGAGATCTGCGGCCAACGCTGCGGAAGTGTGTGCTCAACTTTGAAGAC	1380
Qy	1381	GATTGCTGACGACCAATCATCTCAAGAAATCCCAACGAGATTAATAAATAGGCTAA	1434
Db	1381	GATTGCTGACGACCAATCATCTCAAGAAATCCCAACGAGATTAATAAATAGGCTAA	1434

RESULT 4  
US-10-282-122A-6430  
; Sequence 6430, Application US/10282122A  
; Publication No. US20040029129A1

```

; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert

```

```

? APPLICANT: Ohlsen, Kari
? APPLICANT: Zvekind, Judith
? APPLICANT: Wall, Daniel
? APPLICANT: Trawick, John
? APPLICANT: Carr, Grant
? APPLICANT: Yamamoto, Robert
? APPLICANT: Forsyth, R.
? APPLICANT: Xu, H.
? TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
? FILE REFERENCE: EITRA.034A
? CURRENT APPLICATION NUMBER: US/10/282,122A
? PRIOR FILING DATE: 2003-02-20
? PRIOR APPLICATION NUMBER: 60/151,078
? PRIOR FILING DATE: 2000-03-21
? PRIOR APPLICATION NUMBER: 60/206,848
? PRIOR FILING DATE: 2000-05-23
? PRIOR APPLICATION NUMBER: 60/207,727
? PRIOR FILING DATE: 2000-05-26
? PRIOR APPLICATION NUMBER: 60/230,335
? PRIOR FILING DATE: 2000-09-06
? PRIOR APPLICATION NUMBER: 60/230,347
? PRIOR FILING DATE: 2000-09-09
? PRIOR APPLICATION NUMBER: 60/242,578
? PRIOR FILING DATE: 2000-10-23
? PRIOR APPLICATION NUMBER: 60/253,625
? PRIOR FILING DATE: 2000-11-27
? PRIOR APPLICATION NUMBER: 60/257,931
? PRIOR FILING DATE: 2000-12-22
? PRIOR APPLICATION NUMBER: 60/267,636
? PRIOR FILING DATE: 2001-02-09
? PRIOR APPLICATION NUMBER: 60/269,308
? PRIOR FILING DATE: 2001-02-16
? Remaining Prior Application data removed - See File Wrapper or PALM.
? NUMBER OF SEQ ID NOS: 78614
? SOFTWARE: PatentIn version 3.1
? SEQ ID NO 6430
? LENGTH: 1434
? TYPE: DNA
? ORGANISM: Escherichia coli
US-10-282-122A-6430

Query Match      100.0%; Score 1434; DB 7; Length 1434;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 1434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

ORGANISM: Escherichia coli  
US-10-282-122A-6430

Query Match	Score	DB	Length
100.0%	1434	7	1434

Matches 1434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGAAAGTAAAGCGTCGACAGATTGGAACGTGCACAGATGATGCGCTTGCGAATGATGATG	60
Db	1	ATGAAAGTAAAGCGTCGACAGATTGGAACGTGCACAGATGATGCGCTTGCGAATGATGATG	60
QY	61	CTGATCGTTACTGTGTAACGGCCCAACCAAGTCGTACTCGCCGAAAGCCCGGTC	120
Db	61	CTGATCGTTACTGTGTAACGGCCCAACCAAGTCGTACTCGCCGAAAGCCCGGTC	120
QY	121	GTTAAAGTGAATACCATTCGAAGAAAGTCCTCGGAGCGCGCGCTAACGTGGCGAATGAATATC	180
Db	121	GTTAAAGTGAATACCATTCGAAGAAAGTCCTCGGAGCGCGCGCTAACGTGGCGAATGAATATC	180
QY	181	GCTTCCTCGGTGCTAATGCACGCGCTGATCGGGTTGACAGGGCATTTGACGATGACGCGC	240
Db	181	GCTTCCTCGGTGCTAATGCACGCGCTGATCGGGTTGACAGGGCATTTGACGATGACGCGC	240
QY	241	GCGCTGAATAATCTCTGGCCGACGTCGACGTCGAAATGCGACTTCGTTCTGTACCGACG	300
Db	241	GCGCTGAATAATCTCTGGCCGACGTCGACGTCGAAATGCGACTTCGTTCTGTACCGACG	300
QY	301	CATCCGACCAATTAACCAATTACGGGTACTTCCCGCAACCAACGCTGATCCGCTGTGAT	360
Db	301	CATCCGACCAATTAACCAATTACGGGTACTTCCCGCAACCAACGCTGATCCGCTGTGAT	360
QY	361	TTTGAAGAAGTTTCGAAGGTGTGATCCGACGCGCTGCACGAGCGGATTAATCAGGCG	420
Db	361	TTTGAAGAAGTTTCGAAGGTGTGATCCGACGCGCTGCACGAGCGGATTAATCAGGCG	420



421 CTGAGTTGATTTGGCGCTGCTGTTCTGACTACGCAAGGTCGCTGGCAAGGTA 480  
421 CTGAGTTGATTTGGCGCTGCTGTTCTGACTACGCAAGGTCGCTGGCAAGGTA 480  
481 CAGAGATGATTCACACTGCGCGCTGTAAGCGGCTGTCGCTGATTCACAAAGT 540  
481 CAGAGATGATTCACACTGCGCGCTGTAAGCGGCTGTCGCTGATTCACAAAGT 540  
541 ACCGATTTTGAAGCGCTACCGCGCGCTACGCTGTTAAACCGGATCTCTCGAATTTGAA 600  
541 ACCGATTTTGAAGCGCTACCGCGCGCTACGCTGTTAAACCGGATCTCTCGAATTTGAA 600  
601 GCTGTTCGCTAAATGTAAGACCGAAGAAAGATGTTGAGCGCGGATGAACTGAT 660  
601 GCTGTTCGCTAAATGTAAGACCGAAGAAAGATGTTGAGCGCGGATGAACTGAT 660  
661 GCCGATTACGAACTCTCGCTCTGTTAGTGAACCGCTTCGAAACGAGGATGTCGCTG 720  
661 GCCGATTACGAACTCTCGCTCTGTTAGTGAACCGCTTCGAAACGAGGATGTCGCTG 720  
721 CAACCGGATTAAGCGCGCTGCTGATATGCAACCGAAGCGCAAGATGATGACGTTAC 780  
721 CAACCGGATTAAGCGCGCTGCTGATATGCAACCGAAGCGCAAGATGATGACGTTAC 780  
781 GGTGCGGCGCAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840  
781 GGTGCGGCGCAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840  
841 GAAAGAGCTGCTCTTCTTTCGCAATGCGCGCTGCTGCTGCTGCTGCTGCTGCTG 900  
841 GAAAGAGCTGCTCTTCTTTCGCAATGCGCGCTGCTGCTGCTGCTGCTGCTGCTG 900  
901 TCACAGGTTTGGCGGATGAGCTGGAATGCTGATGCTGAGCTGCAAGATGACGCTTT 960  
901 TCACAGGTTTGGCGGATGAGCTGGAATGCTGATGCTGAGCTGCAAGATGACGCTTT 960  
961 GCGCTGATGACCGAAGAAAGTGAAGCTGCGCTGAGCGCGCGGATGAACTGAGTGA 1020  
961 GCGCTGATGACCGAAGAAAGTGAAGCTGCGCTGAGCGCGCGGATGAACTGAGTGA 1020  
1021 AAGGTGATGATGACCAACGCTGCTTTCGCAATGCTGCAACCGCGGATGCTCTTATCTG 1080  
1021 AAGGTGATGATGACCAACGCTGCTTTCGCAATGCTGCAACCGCGGATGCTCTTATCTG 1080  
1081 GCAAAATGCGCGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140  
1081 GCAAAATGCGCGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140  
1141 AAGGTGATGATGACCAACGCTGCTTTCGCAATGCTGCAACCGCGGATGCTCTTATCTG 1200  
1141 AAGGTGATGATGACCAACGCTGCTTTCGCAATGCTGCAACCGCGGATGCTCTTATCTG 1200  
1201 GCGCTGATGACCGAAGAAAGTGAAGCTGCGCTGAGCGCGCGGATGAACTGAGTGA 1260  
1201 GCGCTGATGACCGAAGAAAGTGAAGCTGCGCTGAGCGCGCGGATGAACTGAGTGA 1260  
1261 ATCCCGGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320  
1261 ATCCCGGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320  
1321 GCGCGGATGATGACCGAAGAAAGTGAAGCTGCGCTGAGCGCGCGGATGAACTGAGTGA 1380  
1321 GCGCGGATGATGACCGAAGAAAGTGAAGCTGCGCTGAGCGCGCGGATGAACTGAGTGA 1380  
1381 GGTGCTGCAAGCAATCATCATCAAGAAAGTTCACAGATTAAGAAAGCTAA 1434  
1381 GGTGCTGCAAGCAATCATCATCAAGAAAGTTCACAGATTAAGAAAGCTAA 1434

RESULT 5  
US-10-771-241-165  
; Sequence 165, Application US/10771241  
; Publication No. US20040241715A1

GENERAL INFORMATION:  
APPLICANT: Zyklich, Judith  
APPLICANT: Forysch, R. Allyn  
TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN  
FILE REFERENCE: ELITRA.001C  
CURRENT APPLICATION NUMBER: US/10/771,241  
CURRENT FILING DATE: 2004-02-03  
PRIOR APPLICATION NUMBER: 09/492,709  
PRIOR FILING DATE: 2000-01-27  
PRIOR APPLICATION NUMBER: 60/117,405  
NUMBER OF SEQ ID NOS: 485  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 165  
LENGTH: 1434  
TYPE: DNA  
ORGANISM: E. coli  
US-10-771-241-165

Query Match 100.0%; Score 1434; DB 8; Length 1434;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGAAAGTAAAGCTGCGCAAGTTTGAACGTCAGAGATGATGATGATGATGATGATGATG 60  
1 ATGAAAGTAAAGCTGCGCAAGTTTGAACGTCAGAGATGATGATGATGATGATGATGATG 60  
61 CTGAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 120  
61 CTGAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 120  
121 GTTAAAGTAAATCAATGCAAGAAAGTCCGCGCGCGCTGATGATGATGATGATGATG 180  
121 GTTAAAGTAAATCAATGCAAGAAAGTCCGCGCGCGCTGATGATGATGATGATGATG 180  
181 GCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240  
181 GCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240  
241 GCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300  
241 GCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300  
301 CATCCGACATTAACCAATTAAGGCTGCTTCCCGCAACCAAGCTGATGATGATGATGATG 360  
301 CATCCGACATTAACCAATTAAGGCTGCTTCCCGCAACCAAGCTGATGATGATGATGATG 360  
361 TTTGAAGAAAGTTCGAAGGCTGATGATGATGATGATGATGATGATGATGATGATGATG 420  
361 TTTGAAGAAAGTTCGAAGGCTGATGATGATGATGATGATGATGATGATGATGATGATG 420  
421 CTGAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 480  
421 CTGAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 480  
481 CAGCAGATGATTCACACTGCGCGCTGTAAGCGGCTGTCGCTGATTCACAAAGT 540  
481 CAGCAGATGATTCACACTGCGCGCTGTAAGCGGCTGTCGCTGATTCACAAAGT 540  
541 ACCGATTTTGAAGCGCTACCGCGCGCTACGCTGTTAAACCGGATCTCTCGAATTTGAA 600  
541 ACCGATTTTGAAGCGCTACCGCGCGCTACGCTGTTAAACCGGATCTCTCGAATTTGAA 600  
601 GCTGTTCGCTAAATGTAAGACCGAAGAAAGATGTTGAGCGCGGATGAACTGAT 660  
601 GCTGTTCGCTAAATGTAAGACCGAAGAAAGATGTTGAGCGCGGATGAACTGAT 660  
661 GCCGATTACGAACTCTCGCTCTGTTAGTGAACCGCTTCGAAACGAGGATGTCGCTG 720  
661 GCCGATTACGAACTCTCGCTCTGTTAGTGAACCGCTTCGAAACGAGGATGTCGCTG 720  
721 CAACCGGATTAAGCGCGCTGCTGATATGCAACCGAAGCGCAAGATGATGACGTTAC 780



QY 661 GCCGATTACGAACTCTGGCTCTGTAGTACCCGTTCCGACAGGATATGCTGCTG 720  
 DB 661 GCCGATTACGAACTCTGGCTCTGTAGTACCCGTTCCGACAGGATATGCTGCTG 720  
 QY 721 CAACCGGATTAAGCGCGCTGATATGCAACCCAGCGGAGAGATGATGACGTTACC 780  
 DB 721 CAACCGGATTAAGCGCGCTGATATGCAACCCAGCGGAGAGATGATGACGTTACC 780  
 QY 781 GGTGCGGCGCAACCGGATTTGGGCTCTGCGCGCAACCGCTGAGGAGGATTAATGCTG 840  
 DB 781 GGTGCGGCGCAACCGGATTTGGGCTCTGCGCGCAACCGCTGAGGAGGATTAATGCTG 840  
 QY 841 GAAGAACCTGCTCTTCTTGGCAATGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTG 900  
 DB 841 GAAGAACCTGCTCTTCTTGGCAATGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTG 900  
 QY 901 TCAACCGTTTCGCAATCGAGCTGGAACCGCGGTGCGCGGTGCGGTGCGGTGCTT 960  
 DB 901 TCAACCGTTTCGCAATCGAGCTGGAACCGCGGTGCGCGGTGCGGTGCGGTGCTT 960  
 QY 961 GCGGTGATGACCGAAGAGAACTGAACTGCGCTGACCGCGCTGACCGCGCTGACCG 1020  
 DB 961 GCGGTGATGACCGAAGAGAACTGAACTGCGCTGACCGCGCTGACCGCGCTGACCG 1020  
 QY 1021 AAAGTGTGATGACCAACGAGTCTTGTGACATCTGACAGCGCGGACGCTCTTATCTG 1080  
 DB 1021 AAAGTGTGATGACCAACGAGTCTTGTGACATCTGACAGCGCGGACGCTCTTATCTG 1080  
 QY 1081 GCAAAATGCGCGCAAGCTGAGTGTGATGATGATGATGATGATGATGATGATGATG 1140  
 DB 1081 GCAAAATGCGCGCAAGCTGAGTGTGATGATGATGATGATGATGATGATGATGATG 1140  
 QY 1141 AAAGGCTGAAAGGAGGATTCGCGCGCGCTGAAACCACTGCAACGCGCTGATGATG 1200  
 DB 1141 AAAGGCTGAAAGGAGGATTCGCGCGCGCTGAAACCACTGCAACGCGCTGATGATG 1200  
 QY 1201 GCGGCACTGAAAGGAGGATTCGCGCGCGCTGAAACCACTGCAACGCGCTGATGATG 1260  
 DB 1201 GCGGCACTGAAAGGAGGATTCGCGCGCGCTGAAACCACTGCAACGCGCTGATGATG 1260  
 QY 1261 ATGCGCGGATCTTGGCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320  
 DB 1261 ATGCGCGGATCTTGGCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320  
 QY 1321 GCGGCGGATTAAGAGTCTGCGCGCAACCGGTGCGGCAATGCTGCTGCTGCTGCTG 1380  
 DB 1321 GCGGCGGATTAAGAGTCTGCGCGCAACCGGTGCGGCAATGCTGCTGCTGCTGCTG 1380  
 QY 1381 GGTGCTGCAACCAATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 1423  
 DB 1381 GGTGCTGCAACCAATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 1423

PRIOR APPLICATION NUMBER: 60/206, 848  
 PRIOR FILING DATE: 2000-05-23  
 PRIOR APPLICATION NUMBER: 60/207, 727  
 PRIOR FILING DATE: 2000-05-26  
 PRIOR APPLICATION NUMBER: 60/242, 578  
 PRIOR FILING DATE: 2000-10-23  
 PRIOR APPLICATION NUMBER: 60/253, 625  
 PRIOR FILING DATE: 2000-11-27  
 PRIOR APPLICATION NUMBER: 60/257, 931  
 PRIOR FILING DATE: 2000-12-22  
 PRIOR APPLICATION NUMBER: 60/269, 308  
 PRIOR FILING DATE: 2001-02-16  
 NUMBER OF SEQ ID NOS: 14110  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 9661  
 LENGTH: 1434  
 TYPE: DNA  
 ORGANISM: *Salmonella typhi*  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (1)...(1434)  
 US-09-815-242-9661  
 Query Match 70.3%; Score 1008.6; DB 3; Length 1434;  
 Best Local Similarity 81.8%; Pred. No. 0;  
 Matches 1164; Conservative 0; Mismatches 259; Indels 0; Gaps 0;

QY 1 ATGAAAGTAAACGCTGCGCAAGATTGAACTGCGAGAGTATGATGATGATGATGATG 60  
 DB 1 ATGAAAGTAAACGCTGCGCAAGATTGAACTGCGAGAGTATGATGATGATGATGATG 60  
 QY 61 CTGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 120  
 DB 61 CTGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 120  
 QY 121 GTTAAAGTAAATCATCATCAAGAAAGCTGCGGCGCGCGCTGAAACGCTGCGATGATG 180  
 DB 121 GTTAAAGTAAATCATCATCAAGAAAGCTGCGGCGCGCGCTGAAACGCTGCGATGATG 180  
 QY 122 GTTAAAGTAAATCATCATCAAGAAAGCTGCGGCGCGCGCTGAAACGCTGCGATGATG 180  
 DB 122 GTTAAAGTAAATCATCATCAAGAAAGCTGCGGCGCGCGCTGAAACGCTGCGATGATG 180  
 QY 181 GCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240  
 DB 181 GCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240  
 QY 181 GCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240  
 DB 181 GCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240  
 QY 241 GCGCTGATGAAATCTGCGCGCAAGCTGCAACGTAATGCAATGCAATGCAATGCAATG 300  
 DB 241 GCGCTGATGAAATCTGCGCGCAAGCTGCAACGTAATGCAATGCAATGCAATGCAATG 300  
 QY 301 CATCGCAATTAACCAATTAACGCGTACTTCCCGCAACCAACGCTGATCCGCTGAT 360  
 DB 301 CATCGCAATTAACCAATTAACGCGTACTTCCCGCAACCAACGCTGATCCGCTGAT 360  
 QY 361 TTTGAAGAGGTTTGAAGGTTGATGATGCGGCGCGCTGCAAGCGGATTAATGCGGCG 420  
 DB 361 TTTGAAGAGGTTTGAAGGTTGATGATGCGGCGCGCTGCAAGCGGATTAATGCGGCG 420  
 QY 421 CTGAGTTGATGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480  
 DB 421 CTGAGTTGATGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480  
 QY 481 CAGCAGATGATCAACTGCGCGCTGAAAGCGGCTGCTGCGGCTGCTGCTGCTGCTGCTG 540  
 DB 481 CAGCAGATGATCAACTGCGCGCTGAAAGCGGCTGCTGCGGCTGCTGCTGCTGCTGCTG 540  
 QY 541 ACCGATTTGAGCGCTACCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600  
 DB 541 ACCGATTTGAGCGCTACCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600  
 QY 601 GCGGTGCGGATTAAGTAAAGCGGAGAGATTGTTGAGGCGGCGGATGAACTGATG 660  
 DB 601 GCGGTGCGGATTAAGTAAAGCGGAGAGATTGTTGAGGCGGCGGATGAACTGATG 660  
 QY 661 GCCGATTACGAACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720  
 DB 661 GCCGATTACGAACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720

D6	661	GCCGATTACGACCTTTCGCGCTGTGGATCGACGCGTTCCGAAACAGGGAATGACGCTGCTG	720
QY	721	CACCCGGGTAAAGCCCGCTGATATGCGCAACCCAAAGCGCAGAAAGTATGACGTTACC	780
D6	721	CAACCGAATPAAAGCCCGCTACTATATGCGCAGCGAGGGCAGGAAATTATGATGTTAAC	780
QY	781	GGTGCGGGCGACACGAGTATGAGCGCTCCGTGGCGGCAACGCTGCGCAGCGGGTAAATTGCTG	840
D6	781	GGTGCGGGCGATACGAGTATCGGCTGCTGGCGGCGACGCGCTGGCGGGAAATACCTTG	840
QY	841	GAAAGACCTGCTTCTTTGCCAATGCGCGGCTGGCGTGGATGGTTCGCGAAACTGCGAAC	900
D6	841	GAGAGAGGGTGTATTTGCCAATGCGCGGCGGGCGGTATGATGTAATCTGGGACG	900
QY	901	TCCACGGTTTGCGCGGATGACGCTGGAAATGCTGTGACGTGCAATACAGGCTTT	960
D6	901	TCAACGGTTTCCCTTATGAGCTGGAAACGCAAGTCCGCGACCGCGGATACCGGCTTC	960
QY	961	GCGGTGATGACCGGAAGAGAACTGAAGCTGGCCGTAAAGCGGCAACGCGCTGTAAC	1020
D6	961	GCGGTGATGACCGGAAGAGAACTGTAAGACAGGCGCGTGCACGCGCGTAACTGGCGAG	1020
QY	1021	AAAGGTGATGACCAACGAGTGTCTTTGACATCTCGACAGCGCGGCAAGTCTTATCTG	1080
D6	1021	AAAGGTGATGACCAACGAGGCTTTTCAATATCTGCAAGCGGGCAAGTCTTATCTG	1080
QY	1081	GCAATGCGCCGCAACGCTGGTGAACCGCTTATGTTGGCTCAACAGGATGCTTCAAC	1140
D6	1081	GCAATGCGCCGCAACCTGGGCGACCGCTGATGTTGGCGTCAATATGACGCTCGACT	1140
QY	1141	AAACGGCTGAAAGGGGATTTCCCGCCCGGTAAACCACTCGAACAGCTGATATGTGCTG	1200
D6	1141	AAACGCTGAAAGGGGAAAGCGCTGCCGTATATCGCTCGAACAGCTGATATGTGCTG	1200
QY	1201	GCGCGACTGGAAGCGGTGCACTGGGTAGTGTGTTGAAAGAGACACGCGCAGCGCTTG	1260
D6	1201	GCGCGCGTGGAGCTCGGTGCACTGGGTGTCTCTTTTGAAGAGATACGCGCAGCACTG	1260
QY	1261	ATCGCCGGGATCTTGCCAGATCTGCTGTGTAAGGCGCGCATATPAAACGAAGAGATT	1320
D6	1261	ATTCGCGGTTATTTCTGCGCGGATCTGTGTGTTAAAGCGCGCGCATPAAACCGGAAGAGT	1320
QY	1321	GCGGGAGTAAAGAGTCTGGGCCCAAGGTGGCGAAGTGTGTGCTCACTTTGAAGAC	1380
D6	1321	GCGGGCAACGAAGAGTCTGGGCCCAACGCGCGCGAATGATGATGCTGAACCTTGAGAT	1380
QY	1381	GTTTGCTGAGCAACCAATCATCTAAGAAAGATCCAAACAGATTA	1423
D6	1381	GGTGTTCACGACCAATATCATCAAAAGAGATCCAGACCGAGA	1423

```

RESULT 8
US-10-282-122A-39294
; Sequence 39294, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zykkind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
;

```

```

; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 39294
; LENGTH: 1434
; TYPE: DNA
; ORGANISM: Salmonella typhi
US-10-282-122A-39294

Query Match      70.3%; Score 1008.6; DB 7; Length 1434;
Best Local Similarity 81.8%; Pred. No. 0;
Matches 1164; Conservative 0; Mismatches 259; Indels 0; Gaps 0;

QY      1 ATGAAAGTAAACCGCTGCGCAGAGTTTGAACGTGCAGAGATGATGCTGTGATGTATG 60
DB      1 ATGAAAGTAAATCTGCGCAGCTTTGAACGTGCAGGCGTCATGCTGTGTGATGTAATG 60

QY      61 CTGATGCTGTATCTGTGTATGCGGCCCAACAGTGTATCTGCGCGGAAGCGCGGTGCGCGT 120
DB      61 CTGATGCTGTATGTGTATGAGGCCCACTTGCCGTATTTACCGGAAGCCCGGTGCGCGT 120

QY      121 GTTAAAGGTAATACCATCATGAAGAAAGTCGCGCGCGCGCTTAAAGTGCAGTAATATC 180
DB      121 GTTAAAGGTAATACCGTTGAGAGAAAGTCGCGCGCGCGCGGAGAACGTGGCATGAACATT 180

QY      181 GCTTCTCTCGGTGCTAATGCAAGCCTGTGCGGTTGACGGGCATTGACGATGCAAGCGCGC 240
DB      181 GCGTCTCTGGAGCGAAGCGCCGCTGTGGTGGCTGACGAGGGTATGATGAAGCGCGCGC 240

QY      241 GCGCTGAGTAATCTCTGCGCGCAGAGTCAAGCTCAAAATGCGACTTCGTTCTGTACCGAG 300
DB      241 GCGCTGAGCAAAACGTGCGCGAGAGTCAATGTGAAGTGCAGCTTCGTTCTGTGCGAAG 300

QY      301 CATCGACCATTAACAATAATTACGGGTAATCTCCGCAACAACAAGCTGATCGTCTGGAT 360
DB      301 CATCGACGATTAACAATCTCGCGTACTATCAGTAATCAGACGCTCATTCGTTGAT 360

QY      361 TTGGAAGAGGTTTGAAGTGTGATCCGACGCGCTGCAAGACGAGATTAAACAGCG 420
DB      361 TTGGAAGAGGCTTTGAGGGGCGTGACCCGACGCGGTTGATGACGATCAACAGCG 420

QY      421 CTGAGTTGATTTGCGCGCGCTGTGCTTTCTGACTACGCAAAAGTGCCTGCAAGCGTA 480
DB      421 CTGGGATGATCGCGCGCTGTGTATGTGCTCAGTTATGCAAAAGGCGCTTACCAAGCGTG 480

QY      481 CAGCAGATGATCCAATCGCGCGGTAAAGCGGGGTGTCCGGGATGATGATCAAAAGGT 540
DB      481 CAGACTATGATTTCCCTTACGCGCGCAAGCGGGCGGTGCTCATGATCCAAAGGA 540

QY      541 ACCGATTTTGAAGCGCTACCGCGCGCTACGCTGTAAACGCGAAATCTCTCGGAATTTGAA 600
DB      541 ACGATTTTGAACGTATACCGCGCGCGCAAGCTGTGTCGACGCAAAACCTTTCTGAATTCGAG 600

QY      601 GCTGTGTGCTGCTAAATGTAAAGCAAGAGATGTTGTTAGCCGCGCGCATGAACACTGATT 660

```

Dp	601	GGGGTGGCGGGGAATGTAAAGCCAAAGCAACGTGTTGAACCGCGCATGAAGCTCATTT	660
Qy	661	GCCGATTACGAACCTTCGGCTCTGTTAAGTACCCGTTCCGAACAGGGATATGTCGTGCTG	720
Dp	661	GCGGATTACGACTTTCCGCGCTGTGTGGTCAACGCGTCCGAACAGGGAATGACGCTGCTG	720
Qy	721	CAACCGGGAAAGCCCGCTGCATATGCGCAACCCAAAGGCAAGAAAGTGTATGACCTTACC	780
Dp	721	CAACCGAATTAAGCCCGCTCATATATGCCAAGCAGCGCGCAAGAAAGTTATGATGTATACC	780
Qy	781	GGTGCGGGCGACACCGGTGATTGGCTCTCGTGGCGGAACGCTGGCACGGGTAAATTCGCTG	840
Dp	781	GGTGCGGGCGAATACGTGTATCGGCGTGTGCGGGCGAACGCTGGCGGGAAATACCTG	840
Qy	841	GAAGAGCCTGCTTTCTTTGCCAATGCGGCGGCTGGCGTGGTGTCTGGCAAACTGGGAACC	900
Dp	841	GAGGAGGCGTGTATTATTTGCGCAATGCGGCGGCGGGCGTGTGTGTATTAATTCGGGACG	900
Qy	901	TCCAGGGTTTGGCCGATCGAGCTGGAAAAATGCTGTACGTGACGTGCAATTCAGGCTTT	960
Dp	901	TCCAGGGTTTCCCTTATTGAGCTGGAAAAACGAGTGGCGGGACGGCGGATACCGGCTTC	960
Qy	961	GCGGTGATGACCGGAAGAGAACTGAAGCTGGCGCGTACGCGGCAAGCGTAAACGTGGTGA	1020
Dp	961	GCGGTCAATGACCGGAAGAGAACTTGAAGACAGGCGCGTGCACGCGCGGTATAGCGTGGCAG	1020
Qy	1021	AAAGTGTGATGACCAACCGGTGCTTTTGAATCTCTGCAACGCGGGCAAGCTCTTATCTG	1080
Dp	1021	AAAGGTGTCATGACCAACGCGCTTTTCAATTTATTCGACAGCGGGCAAGCTCTTATCTG	1080
Qy	1081	GCAATGCCCCGCACTGGGTATACCGCTTATTTGTTCGCTCAACAGAGATGCTTCACAC	1140

```

/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
/ FILE REFERENCE: ELITRA.034A
/ CURRENT APPLICATION NUMBER: US/10/282,122A
/ PRIOR FILING DATE: 2003-02-20
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 24093
/ LENGTH: 1431
/ TYPE: DNA
/ ORGANISM: Klebsiella pneumoniae
US-10-282-122A-24093

```

```

Qy 541 ACCGATTTTGACGCTACCGGCGCTACGCTGTAAACCGCAATCTCTCGAATTTGAA 600
Db 541 ACCGATTTTGACGCTACCGGCGCTACCGGCGCTACCGGCGCTCTCTGAGTTTGA 600
Qy 601 GCTGTGTGCTTAATGTAAGACCGAAGAAAGATTTGTAGCGCGGCTAATGAACTATT 660
Db 601 GCGGTGTGTGCTTAATGTAAGACCGAAGAAAGATTTGTAGCGCGGCTAATGAACTATT 660
Qy 661 GCCGATTAACGAATCTGCGCTCTGTAGTACCGGCTCCGACAGGATATGTCGCTG 720
Db 661 GCCGATTAACGAATCTGCGCTCTGTAGTACCGGCTCCGACAGGATATGTCGCTG 720
Qy 721 CAACCGGATTAACGCGCTGCTATATGCAACCAAGCGACGAAGTATGATGCTTACC 780
Db 721 CAACCGGATTAACGCGCTGCTATATGCAACCAAGCGACGAAGTATGATGCTTACC 780
Qy 781 GGTGCGGCGGACAGGATGATTTGGGCTCTGTGCGGCAACGCTGGAAGGGGATTTGCTG 840
Db 781 GGTGCGGCGGACAGGATGATTTGGGCTCTGTGCGGCAACGCTGGAAGGGGATTTGCTG 840
Qy 841 GAAGAGGCTGCTCTTTTGCATATGCGGCGGCTGCGGCTGCTGCGCAAACTGGAACC 900
Db 841 GAAGAGGCTGCTCTTTTGCATATGCGGCGGCTGCGGCTGCTGCGCAAACTGGAACC 900
Qy 901 TCACCGGTTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
Db 901 TCACCGGTTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
Qy 961 GCGCTGATGACCGAAGGAACTGAAGCTGCGGCTGCGGCGGCTGCGGCGGCTGCGG 1020
Db 961 GCGCTGATGACCGAAGGAACTGAAGCTGCGGCTGCGGCGGCTGCGGCGGCTGCGG 1020
Qy 1021 AAAGTGTGATGACCAACGCTGCTTGTGACATCTGCAACGCGGCGGCTGCTTATCTG 1080
Db 1021 AAAGTGTGATGACCAACGCTGCTTGTGACATCTGCAACGCGGCGGCTGCTTATCTG 1080
Qy 1081 GCAATGCGCGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
Db 1081 GCAATGCGCGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
Qy 1141 AAAGCTGTAAGAGGAGTTCGCGGCGGCTGTAACCGGCTGTAAGAGGAGTTCGCTG 1200
Db 1141 AAAGCTGTAAGAGGAGTTCGCGGCGGCTGTAACCGGCTGTAAGAGGAGTTCGCTG 1200
Qy 1201 GCGGCACTGGAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
Db 1201 GCGGCACTGGAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
Qy 1261 ATGCGCGGATCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
Db 1261 ATGCGCGGATCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
Qy 1321 GCGGCGGATTAAGAGTTCGCGGCGGCTGTAACCGGCTGTAAGAGTTCGCTGCTG 1380
Db 1321 GCGGCGGATTAAGAGTTCGCGGCGGCTGTAACCGGCTGTAAGAGTTCGCTGCTG 1380
Qy 1381 GGTGCTGCAAGCAACATCATCAAGAGATCCCAACAGATA 1423
Db 1381 GGTGCTGCAAGCAACATCATCAAGAGATCCCAACAGATA 1423

```

## RESULT 10

US-10-282-122A-36984

```

; Sequence 36984, Application US/10282122A
; Publication No. US20040029129A1

```

## GENERAL INFORMATION:

```

; APPLICANT: Wang, Jiansu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel

```

```

; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OR INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 36984
; LENGTH: 1397
; TYPE: DNA
; ORGANISM: Salmonella paratyphi A
US-10-282-122A-36984

```

```

Query Match 65.2%; Score 935.2; DB 7; Length 1397;
Best Local Similarity 81.1%; Pred. No. 6.1e-298;
Matches 1137; Conservative 0; Mismatches 258; Indels 7; Gaps 4;

```

```

Qy 22 TTGAAAGTGCAGAGTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 81
Db 1 TTGAAAGTGCAGAGTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 60
Qy 82 CCACAGAGTATCTGCGCGGAAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 141
Db 61 CCACAGTTCGCTATTTCAACGAAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
Qy 142 GAAAGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 201
Db 121 GAAAGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
Qy 202 GCGCTGCTGCGGTTGACGCGGATTTGACAGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 261
Db 181 GCGCTGCTGCGGTTGACGCGGATTTGACAGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 238
Qy 262 GAGTCAAGTCAAAAGCGAGCTTGTCTGTAACGAGCGATCGACATTAACCAATTA 321
Db 239 GAGTCAATGTGAAGTGCAGCTTGTCTGTAACGAGCGATCGACATTAACCAATTA 298
Qy 322 CGGATCTTCCCGCAACCAACAGCTGATCCGTCTGATTTTGAAGAGTTTGAAGGT 381
Db 299 CGGATCTTCAAGTATCAAGTATCAAGCTGATCCGTCTGATTTTGAAGAGTTTGAAGGT 358
Qy 382 GTTGATCCGAGCGCGCTGCAAGAGCGGATTAATCAAGCGCTGATTTGAGTTGCGCGCTG 441
Db 359 GTGAGCCGCGAGCGGTTGCAAGAGCGGATTAATCAAGCGCGCTGATTTGAGTTGCGCGCTG 418
Qy 442 GTGCTTTCGATCAAGCGCAAGAGTTCGCGCGGATTAATCAAGCGCGCTGATTTGAGTTGCGCGCTG 501
Db 419 GTATGTCGATTAAGCAAGAGCGGCTGCAAGCGGATTAATCAAGCGCGCTGATTTGAGTTGCGCGCTG 478

```

QY 502 CGTAAAGGGGTTCCGGTCTGATGATCCAAAGTACCGATTTGAGCCCTACCC 561  
 DB 479 CGCAGGGGGGCGCGGCTGATGATCCGAAAGGAAAGGATTTTGAACGTTACCGC 538  
 QY 562 GGGCGCTACGCTGTTAAGCCGAAATCTCTCGGAATTTGAGCTGTTGCTGTAATGTAG 621  
 DB 539 GGGCGCAAGCTGCTGAGCGCCAAACCTTCTGAAATTCAGAGCGGTTGGGGGAAATGTAA 598  
 QY 622 ACCGAAAGAGATTGTTAGCGCGGCAAGAACTGATTCGATTAAGAACTCTCGGCT 681  
 DB 599 AGCCAAAGCAAGCTGTTGAACGGGCAATGAACTCATTCGATTCGATTCGCGC 658  
 QY 682 CTGTTAGTACCCGTTCCGAAACAGGATATGCTGCTGCAACCGGGTAAAGCCCGCTG 741  
 DB 659 CTGTTGCTGACGCTTCCGAAACAGGAAATGACGCTGCTGCAACCGAAATTAAGCCGCTA 718  
 QY 742 CATATGCCAACCCAAAGGCAAGAAATGATGACCTTACCGGTGCGGCGACACGCTGAT 801  
 DB 719 CATATGCCAACCGAGGCGAGAAATTTATGATTTACCGGTGCGGCGATACGCTGATC 778  
 QY 802 GGGCTCTGGCGGCAAGCTGCGACGCGGATATTCCTGGAAGAGCTCTCTTCTTCC 861  
 DB 779 GGGCTCTGGCGGCGAC-CTGGCGGCGGAAATACCTGGAAGAGCGCTGTTATTTCC 837  
 QY 862 AATGCGGCGGCTGGCGGTGCTGGGAAATCTGGAACCTTCCAGGTTTCCGATCGAG 921  
 DB 838 AATGCGGCGGCGGCGGTGATGATGATTAATCTGGGAGCTGTAACGCTTATTTAG 897  
 QY 922 CTGGAATAATGCTGATGCTGACGATGACGCTTGGCGGTGATGACGGAAGAA 981  
 DB 898 CTGGAATAATGCTGATGCTGACGATGACGCTTGGCGGTGATGACGGAAGAA 954  
 QY 982 CTGGAATGCTGATGCTGACGATGACGCTTGGCGGTGATGACGGAAGAA 1041  
 DB 955 TTGAGAGAGGCGCTGCGCAGCGCGCTGAAAGCTGCGGCAAGAGTGCATGACCAAGCGC 1014  
 QY 1042 GTCTTGAACATCTGCAACGCGGCGACGCTCTTATCTGGGAAATGCGCCCAAGCTGGT 1101  
 DB 1015 GTTTCATATTTCTGCAACGCGGCGACGCTCTTATCTGGGAAATGCGCCCAAGCTGGT 1074  
 QY 1102 GACCGCTGATGATGCTGCGCTGCAACAGCGATGCTCTCAACCAAGCGCTGAAGGAGATTCC 1161  
 DB 1075 GACCGCTGATGATGCTGCGCTGCAACAGCGATGCTCTCAACCAAGCGCTGAAGGAGATTCC 1134  
 QY 1162 GCGCGGCTGAACCACTGCAACAGCGATGATGCTGCTGCGGCGCACTGGAAGCGCTGAC 1221  
 DB 1135 GCTCGGCTGAATCCGCTGCAACAGCGATGATGCTGCTGCGGCGCGCTGAGTCCGTCGAC 1194  
 QY 1222 TGGGTATGCTGCTTGAAGAGGAAACGCGGAGGCTTGAATCGCGGAGATCTTCCAGAT 1281  
 DB 1195 TGAATGCTCTTGAAGAGGAAATGCGCGCAAGCTGATGCTGCGGATTTCTGCGGAT 1254  
 QY 1282 CTGCTGCTGAAGAGCGCGCTGATTAACAGAGAGATGCGCGGAGTGAAGAGATCTGG 1341  
 DB 1255 CTGCTGCTGAAGAGCGCGCTGATTAACAGAGAGATGCGCGGAGTGAAGAGATCTGG 1313  
 QY 1342 GCCAAGCGTGGCGAAGTGTGTCTCAACTTTGAAGAGCGGTGCTGACGACCAATC 1401  
 DB 1314 GCCAAGCGGCGGAGATGATGCTGCACTTGAAGAGTGTGTCTCAAGCAATATC 1373  
 QY 1402 ATCAAGAGATCCAAAGAT 1423  
 DB 1374 ATCAAGAGATCCAAAGAT 1395

APPLICANT: Malone, Cheryl  
 APPLICANT: Haselbeck, Robert  
 APPLICANT: Ohlsen, Kari  
 APPLICANT: Zykkind, Judith  
 APPLICANT: Wall, Daniel  
 APPLICANT: Trawick, John  
 APPLICANT: Carr, Grant  
 APPLICANT: Yamamoto, Robert  
 APPLICANT: Forsyth, R.  
 TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
 FILE REFERENCE: ELITRA.034A  
 CURRENT APPLICATION NUMBER: US/10/282,122A  
 PRIOR FILING DATE: 2003-02-20  
 PRIOR FILING DATE: 60/191,078  
 PRIOR FILING DATE: 2000-03-21  
 PRIOR FILING DATE: 60/206,848  
 PRIOR FILING DATE: 2000-05-23  
 PRIOR FILING DATE: 60/207,727  
 PRIOR FILING DATE: 2000-05-26  
 PRIOR FILING DATE: 60/230,335  
 PRIOR FILING DATE: 2000-09-06  
 PRIOR FILING DATE: 60/230,347  
 PRIOR FILING DATE: 2000-09-09  
 PRIOR FILING DATE: 60/242,578  
 PRIOR FILING DATE: 2000-10-23  
 PRIOR FILING DATE: 60/253,625  
 PRIOR FILING DATE: 2000-11-27  
 PRIOR FILING DATE: 60/257,931  
 PRIOR FILING DATE: 2000-12-22  
 PRIOR FILING DATE: 60/267,636  
 PRIOR FILING DATE: 2001-02-09  
 PRIOR FILING DATE: 60/269,308  
 PRIOR FILING DATE: 2001-02-16  
 Remaining Prior Application data removed - See File Wrapper or PALM.  
 NUMBER OF SEQ ID NOS: 78614  
 SOFTWARE: Patent version 3.1  
 SEQ ID NO 41981  
 LENGTH: 1431  
 TYPE: DNA  
 ORGANISM: Yersinia pestis  
 US-10-282-122A-41981

Query Match 61.3%; Score 879.2; DB 7; Length 1431;  
 Best Local Similarity 76.2%; Pred. No. 2.1e-279;  
 Matches 1082; Conservative 0; Mismatches 338; Indels 0; Gaps 0;

QY 1 ATGAAGTAAAGCTGCGCAAGTTTGAACGTGACAGAGTATGCTGCTGATGATG 60  
 DB 1 ATGAAGTAAAGCTGCGCAAGTTTGAACGTGACAGAGTATGCTGCTGATGATG 60  
 QY 61 CTGATGCTTACTGTGACGCGCCACAGCTGATCTGCGCGGAGCGCGGCTGG 120  
 DB 61 TTGACCGCTACTGTGACGCGCCACAGCTGATCTGCGCGGAGCGCGGCTGG 120  
 QY 121 GTTAAGTAAATCTATCGAAGAACTCCGGGCGCGCTTAACTGCGGATTAATC 180  
 DB 121 GTTAAGTAAATCTATCGAAGAACTCCGGGCGCGCTTAACTGCGGATTAATC 180  
 QY 181 GCTTCTCTGCGGCTATGACAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240  
 DB 181 GCTTCTCTGCGGCTATGACAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240  
 QY 241 GCGCTAGTAAATCTGCGCGCAAGTCAAGTCAATGCGACTTCTGTTCTGACGAGC 300  
 DB 241 GCGCTAGTAAATCTGCGCGCAAGTCAAGTCAATGCGACTTCTGTTCTGACGAGC 300  
 QY 301 CATCCAGCATTAACCAATTAACGCGTATCTCCGCAACCAACAGCTGATCCGCTGAT 360  
 DB 301 CATCCAGCATTAACCAATTAACGCGTATCTCCGCAACCAACAGCTGATCCGCTGAT 360  
 QY 361 TTGAAGAGGTTTCAAGGTTGATCGCAGCGCTGCAAGAGCGGATTAATCAGGCG 420

RESULT 11  
 US-10-282-122A-41981  
 Sequence 41981, Application US/10282122A  
 Publication No. US20040029129A1  
 GENERAL INFORMATION:  
 APPLICANT: Wang, Liangsu  
 APPLICANT: Zamudio, Carlos







[illegible]

```
Db      1981 GGTATTTCACAACTAATATTATTAATGACCATTAATAAGAAA 1421

RESULT 13
US-10-282-122A-22349
; Sequence 22349, Application US/10282122A
; Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See file Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 22349
LENGTH: 1431
TYPE: DNA
ORGANISM: Haemophilus influenzae
US-10-282-122A-22349

Query Match          44.3%; Score 635.2; DB 7; Length 1431;
Best Local Similarity 65.9%; Pred. No. 1.le-198;
Matches 922; Conservative 0; Mismatches 478; Indels 0; Gaps 0;
```

257 TGGCCGACGTACAGTCATATGCACTTCTGTTCTGTATCCGACGATCCGACATTAACA 316  
257 TAGAAAAACAAAATGATTTGTTATTTGTTGATTAAGAACCATTCACATTTTCTA 316  
317 AATTACGGGTACTTCCCGCAACCAACGCTGATCCGCTGATTTTGAAGAAGTTTCG 376  
317 AATTACGATTTTATCTCGTCATCAACGCTGCTCCGCTTGATTTGAAAGAAATTTCA 376  
377 AAGGTTGATCCGACGCGTGCACGAGGGAATTAATCAAGCGCTGATTCGATGGCG 436  
377 ATAAATGATGATTCAGAGATTTATTAAGCAATTAAGAAAGTGCCTGAAAATTAACGTTG 436  
437 CGCTGTGCTTTTCTGACTACCCAAAGGTGCTGCGACGCGTACAGAGATGATCCAC 496  
437 CTTTGATTTCTTCTGATTAAGGCAAGCGCTTAAAGATTTCAAGAAATGATTTCAA 496  
497 TGGCGCGTAAACGGGGTGTCCGGTCTGATTAATCCAAAGTAAAGTATTTGAGCGCT 556  
497 TTGCAACGCAAGCAATGCTGCTGTTGATCGATCCAAAGGAACTGATTTTGAAGCTT 556  
557 ACCGCGGCGCTACGCTGTTAAAGCGCAATCTCTGGAAATTTGAAGCTGTTGCGTAAAT 616  
557 ATCGTGGGCTACATTAATTAAGCAACCAATATGCTGAATTTGAAGCTGTTGGGTAAT 616  
617 GTAAGACCGAAGAAAGATTTGTGACGCGCATGAATGATTTGCCGATTAACGACTCT 676  
617 GCAATACGGAAGAAAGATTTATTAAGAAAGTTTAAATTAATTTCTGATTAATTA 676  
677 CGGCTCTGTATGTAACCCGTTCCGAACAGGGTATGTCGCTCTGCAACCGGCTAAAGCC 736  
677 CCGCATTTTGTGATCGCGTCTGTAAGAAAGCATGACATTAATTAACGCCAATCAAGAC 736  
737 CGCTCATATGACCAACCGCAAGCGGAGAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 796  
737 CTTATCATTTTCCCACTGTTGCAAAAGAGTGTGATGATGACGAGACCGGTACACTG 796  
797 TGATTTGGGCTCTGCGCGCAACGCTGCGAGCGGTTAATTCCTGTAAGAAAGCTGCTTCT 856  
797 TCATTAAGGATTAAGCAACCGCATTAAGAGATGAGCGTTCTTGAAGAAATCTTGTTAAC 856  
857 TTGCAATGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 916  
857 TAGCCAAAGTGTGCGCAAGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 916  
917 TCGAGCTGGAATATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 976  
917 TCGAATTTAAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 976  
977 AGGAATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1036  
977 CAGATTTAAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1036  
1037 ACGGTGCTTTTGAATCTGCAACGCGCGCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1096  
1037 ATGCTGTTTGAATTTTGAATCTGCAACGCGCGCAAGTCTGCTGCTGCTGCTGCTGCTGCTG 1096  
1097 TGGGTAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1156  
1097 TGGGTAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1156  
1157 AAGGTGCGCAATTAATTAATCTGCAACGCGCGCAAGTCTGCTGCTGCTGCTGCTGCTGCTG 1216  
1157 ATTCCCGCCGCGTAAACCACTGCAACGCGCGCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1216  
1217 TCGACTGGGTAGTGTCTGTTGAAGAGCAAGCGCGCAAGTCTGCTGCTGCTGCTGCTGCTGCTG 1276  
1217 TAGACTGGGTAGTGTCTGTTGAAGAGCAAGCGCGCAAGTCTGCTGCTGCTGCTGCTGCTGCTG 1276  
1277 CAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1336  
1277 CAGATCTTTTTGAATCTGCAACGCGCGCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1336  
1337 TCTGGGCAACGCTGCGCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1396

Db 1337 TTTGGGCAACGCTGCGAGTGAAGTCTAACTTTGAAGTGTGTTCAACAA 1396  
Qy 1397 ACATCATCAGAGATCCAA 1416  
Db 1397 ATGTGATGAAAAAATTTAA 1416

RESULT 14  
US-10-329-670-1/c  
; Sequence 1, Application US/10329670  
; Publication No. US20040018503A1  
; GENERAL INFORMATION:  
; APPLICANT: Fleischmann et al.  
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Fragment of Invention: Thereof, and Uses Thereof  
; FILE REFERENCE: P186P1  
; CURRENT APPLICATION NUMBER: US/10/329,670  
; PRIOR FILING DATE: 2002-12-24  
; PRIOR APPLICATION NUMBER: US 09/643,990  
; PRIOR FILING DATE: 2000-08-23  
; PRIOR APPLICATION NUMBER: US 08/487,429  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: US 08/426,787  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 1830121  
; TYPE: DNA  
; ORGANISM: Haemophilus influenzae  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (4747)..(4747)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (9921)..(9921)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (10150)..(10150)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (29298)..(29298)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (36543)..(36543)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (36551)..(36551)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (36636)..(36636)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (40810)..(40810)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (44416)..(44416)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (44905)..(44905)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature

```

1 OTHER INFORMATION: n equals a, t, g or c
2 FEATURE:
3 NAME/KEY: misc_feature
4 LOCATION: (120038) .. (120038)
5 OTHER INFORMATION: n equals a, t, g or c
6 FEATURE:
7 NAME/KEY: misc_feature
8 LOCATION: (121344) .. (121344)
9 OTHER INFORMATION: n equals a, t, g or c
10 FEATURE:
11 NAME/KEY: misc_feature
12 LOCATION: (122167) .. (122167)
13 OTHER INFORMATION: n equals a, t, g or c
14 FEATURE:
15 NAME/KEY: misc_feature
16 LOCATION: (122336) .. (122336)
17 OTHER INFORMATION: n equals a, t, g or c
18 FEATURE:
19 NAME/KEY: misc_feature
20 LOCATION: (131340) .. (131340)
21 OTHER INFORMATION: n equals a, t, g or c
22 FEATURE:
23 NAME/KEY: misc_feature
24 LOCATION: (131360) .. (131360)
25 OTHER INFORMATION: n equals a, t, g or c
26 FEATURE:
27 NAME/KEY: misc_feature
28 LOCATION: (139910) .. (139910)
29 OTHER INFORMATION: n equals a, t, g or c
30 FEATURE:
31 NAME/KEY: misc_feature
32 LOCATION: (140398) .. (140398)
33 OTHER INFORMATION: n equals a, t, g or c
34 FEATURE:
35 NAME/KEY: misc_feature
36 LOCATION: (145058) .. (145058)
37 OTHER INFORMATION: n equals a, t, g or c
38 FEATURE:
39 NAME/KEY: misc_feature
40 LOCATION: (145171) .. (145171)
41 OTHER INFORMATION: n equals a, t, g or c
42 FEATURE:
43 NAME/KEY: misc_feature
44 LOCATION: (145942) .. (145942)
45 OTHER INFORMATION: n equals a, t, g or c
46 FEATURE:
47 NAME/KEY: misc_feature
48 LOCATION: (147197) .. (147197)
49 OTHER INFORMATION: n equals a, t, g or c
50 FEATURE:
51 NAME/KEY: misc_feature
52 LOCATION: (150841) .. (150841)
53 OTHER INFORMATION: n equals a, t, g or c
54 FEATURE:
55 NAME/KEY: misc_feature
56 LOCATION: (152500) .. (152500)
57 OTHER INFORMATION: n equals a, t, g or c
58 FEATURE:
59 NAME/KEY: misc_feature
60 LOCATION: (152530) .. (152530)

```

	Query Match	44.3%	Score 635.2	DB 7	Length 1830121
	Best Local Similarity	65.9%	Pred. No.2.9e-197		
	Matches 922	Conservative 0	Mismatches 478	Indels 0	Gaps 0
OY	17	CAGAGTTGAAACGCGCAGAGTGAATGCTGTGCTGATGATGATCTCGATCGTTACTCGT			76
Db	1596554	CAGAAITTAAGCAACAAAGTACTCGATTATAGCGACGTGATCTCTTGATCGTTATTCGT			1596495

QY 77 ACGGCCCCACGATCTCTGCGGAGAGCGGCGGCGGCTTAAAGTAATCA 136  
Db 1596494 TCGGCGCAACCGATATTTTACACAGAGACGAGTGTTCGTGTACAGAAA 1596435  
QY 137 TCGAAGAACGTCGCGGCGGCGGCTTAAAGTGAATATGCTTCTCGGTCTA 196  
Db 1596434 ATGAAGAACGCGGCGGCGGCTTAAAGTGAATATGCTTCTCGGTCTA 1596375  
QY 197 ATGACGCGGCGGCGGCTTAAAGTGAATATGCTTCTCGGTCTA 256  
Db 1596374 CCGTTCAGTTAATGAGATTTGAACAGATGAACTGCTTCTCGGTCTA 1596315  
QY 257 TGGCGGACGTCACGTCACATGCACTTCTGTAACGACGATCCGACATCA 316  
Db 1596314 TAGAAAAACAAAAATGATTTGTAATTTTGTGATTAAGAAACCATCA 1596255  
QY 317 AATTACGGGATCTTCCCGCAACACAGCTGATCCGCTGATTTGAAAGGTT 376  
Db 1596254 AATTACGGATTTTATCTGTCATCAACAGCTGCTCCGCTTGAATTTGA 1596195  
QY 377 AAGGTTTGAATCCGACGCGCTGACAGCGGATTAATACAGCGCTGATG 436  
Db 1596194 AATAATGATGATTCAGAGATTTATTAAGCAAGTTAAGAAAGTGGTAA 1596135  
QY 437 CGCTGCTGCTTCTGATACGCAAGGTCGCTGCAAGCGTACAGACATGAT 496  
Db 1596134 CTTTGATTTCTTCTGATTAAGCAAGGTCGCTTAAAGATTTCAAGAAAT 1596075  
QY 497 TGGCGGTAAGCGGCGGCTTCCGCTGCTGATTTCCAAAAGTACCGATTT 556  
Db 1596074 TTGACCGCAAGCAAGTGTCTGTTGATGATTCGCAAGGCAATG 1596015  
QY 557 ACGCGGCGCTACGCTGTTAAGCGCAATCTCTGCAATTTGAAGCTTTG 616  
Db 1596014 ATCTGGGCGCTACATTAATGACACCATTAATGCTGATTTGAAG 1595955  
QY 617 GTAAGACGGAAGAGATTTGTAAGCGGCAATGAATGCGATTAACGAC 676  
Db 1595954 GCAATACGGAAGAGATTTGTAAGAGGTTAAATTAATTTGATTAAT 1595895  
QY 677 CGGCTGCTGATGATCCGCTTCCGACAGGATATGCTGCTGCAACGG 736  
Db 1595894 CCGGCTTTGCTGATCGGCTTCTGAAAAGGCAATGATTAATCAAGC 1595835  
QY 737 CGTGCATATGCTCAACCGACGAGAGTGAATGAGCTTAAAGCTGTCG 796  
Db 1595834 CTATATCATTTGCTCAATGCTTGAAGAAAGATTTGATGATGACG 1595775  
QY 797 TGATTTGCGCTCTGCGGCAACGCTGCGAGCGGTAATTCGCTGAA 856  
Db 159774 TCATTAAGGTAATGCAACCGCATTAAGAGATGACGTTCTTGA 1595715  
QY 857 TTGCGCAATGCGGCGGCTGCTGCTGCTGCAAACTCCACGCTTCCG 916  
Db 1595714 TAGCCAAAGTTGCGGCAAGAAATGCTGAGAAATGCGGCTTCA 1595655  
QY 917 TCGAGCTGAAGAAATGCTGATAGTGAAGTGAATGAGCTTGAAGCTG 976  
Db 1595654 TCGAATCTGAAGAAATGCTGATAGTGAAGTGAATGAGCTTGA 1595595  
QY 977 AGGAATGGAAGCTGCGCTGACGCGGCAAGCTGCTGAAAGTGTG 1036  
Db 1595594 CAGAATTTAAAGATGCTGCTGCAACAGCTAAAGCGGCTGAAAGT 1595535  
QY 1037 ACGGTGCTTTGATCTCTGACGCGGCGGCTGCTTATCTGCAAAAT 1096  
Db 159534 ATGCTGTTTGAATTTTGAATTCAGGCGCAATTTCTTATTTGA 1595475  
QY 1097 TGGGTAACGCTGATTTGCTGCGCTCAACAGAGATGCTTCAACAAA 1156  
Db 1595474 TGGGCGATGCTGATTTGCTGCGGTAACAGGAGATTTCTTAA 1595415  
QY 1157 ATTCCCGCCGCTAAACCACTCGAAGAGCTGATGATGCTGCGCG 1216

Db 1595414 AAGTCCCAATTAATTTGAAAGCGATGCGGATGCTGCTTGGCATCCG 1595355  
QY 1217 TCGATGCGGATGCTTTGAAAGACACGCGGAGGCTTATGCGCGGATCT 1276  
Db 1595354 TAGATGCTGCTGCTTCTTCACTGAAGATACACACGCTTAAAT 1595295  
QY 1277 CAGATGCTGCTGAAAGCGGCGGCTATTAACAGAGATGCTGCGGAGT 1336  
Db 1595294 CAGATCTTTATGTCAAAGGCGGCTTACAAACCGAAGATTTG 1595235  
QY 1337 TCGGCGCAACGCTGCGGAGATGCTGCTCACTTTGAAGCGTGT 1396  
Db 1595234 TTGCGGCAACGCTGCGGAGATGTTAAAGTGTGTTAAACAA 1595175  
QY 1397 ACATCATCAAGATCCAA 1416  
Db 1595174 ATGATGATGAAAAATTA 1595155

RESULT 15  
US-10-158-865-1/c  
; Sequence 1, Application US/10158865  
; Publication No. US20040203093A1  
; GENERAL INFORMATION:  
; APPLICANT: Fleischmann et al.  
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome, Fragm  
; TITLE OF INVENTION: Thereof, and Uses Thereof  
; FILE REFERENCE: PB186P2C1D1  
; CURRENT APPLICATION NUMBER: US/10/158, 865  
; CURRENT FILING DATE: 2002-06-03  
; PRIOR APPLICATION NUMBER: US 09/557,884  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: US 08/476,102  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: US 08/426,787  
; PRIOR FILING DATE: 1995-04-21  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 1830121  
; TYPE: DNA  
; ORGANISM: Haemophilus Influenzae  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4747)..(4747)  
; OTHER INFORMATION: n equals a,t,c, or g  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (9921)..(9921)  
; OTHER INFORMATION: n equals a,t,c, or g  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (10150)..(10150)  
; OTHER INFORMATION: n equals a,t,c, or g  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (29298)..(29298)  
; OTHER INFORMATION: n equals a,t,c, or g  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (36543)..(36543)  
; OTHER INFORMATION: n equals a,t,c, or g  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (36551)..(36551)  
; OTHER INFORMATION: n equals a,t,c, or g  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (36636)..(36636)  
; OTHER INFORMATION: n equals a,t,c, or g  
; FEATURE:  
; NAME/KEY: misc feature

```
LOCATION: (40808) .. (40810)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44416) .. (44416)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44905) .. (44905)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44975) .. (44975)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45593) .. (45593)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45732) .. (45732)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (47036) .. (47036)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51334) .. (51334)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51602) .. (51602)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51786) .. (51786)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51805) .. (51805)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (53369) .. (53369)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65309) .. (65309)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65313) .. (65313)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (80024) .. (80024)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (100091) .. (100091)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (102696) .. (102696)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (105121) .. (105121)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (107246) .. (107246)
OTHER INFORMATION: n equals a,t,c, or g

OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (117136) .. (117136)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119750) .. (119750)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119924) .. (119924)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (120038) .. (120038)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (121344) .. (121344)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122167) .. (122167)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122336) .. (122336)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131340) .. (131340)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131360) .. (131360)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (139910) .. (139910)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (140398) .. (140398)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (142750) .. (142750)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145058) .. (145058)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145171) .. (145171)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145942) .. (145942)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (147197) .. (147197)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (150841) .. (150841)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152500) .. (152500)
OTHER INFORMATION: n equals a,t,c, or g
```



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2006, 18:51:39 ; Search time 938 Seconds  
(without alignments)

10188.880 Million cell updates/sec

Title: US-10-520-820-28

Perfect score: 1434  
Sequence: 1 atgaagatgaacgtctgcaga.....aacagataaaaaagctaa 1434

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Genesegq.21:\*  
2: genesegq1980s:\*  
3: genesegq1990s:\*  
4: genesegq2000s:\*  
5: genesegq2001as:\*  
6: genesegq2002as:\*  
7: genesegq2002bs:\*  
8: genesegq2003as:\*  
9: genesegq2003bs:\*  
10: genesegq2003cs:\*  
11: genesegq2003ds:\*  
12: genesegq2004as:\*  
13: genesegq2004bs:\*  
14: genesegq2005s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1434	100.0	1434	3	AA65973
2	1434	100.0	1434	4	AA65973
3	1434	100.0	1434	5	AA65973
4	1434	100.0	1434	6	AA65973
5	1434	100.0	1434	7	AA65973
6	1018.2	71.0	1428	8	ACA32148
7	1011.8	70.6	1500	11	ACH96242
8	1008.6	70.3	1434	4	AA65973
9	1008.6	70.3	1434	4	AA65973
10	1007.2	70.2	1434	8	ACA51424
11	995.2	65.2	1397	8	ACA51424
12	879.2	61.3	1431	8	ACA51424
13	870.6	60.7	1323	11	ACH96312
14	784.8	54.7	110000	10	ACF65388_02
15	784.8	54.7	110000	10	ACF65388_02
16	783.4	54.6	1425	10	ACF71372
17	725	50.5	1422	10	ACA44666
18	724.2	50.5	1476	10	ADP02177
19	635.2	44.3	1431	8	ACA34479

C	20	635.2	44.3	110000	2	AA642063.15	Continuation (16 o
	21	625.6	43.6	349980	13	ADP05377	Ad05377 Haemophil
	22	617.4	43.1	1431	8	ACA33070	ACA33070 Prokaryot
	23	507.8	35.4	707	6	ABQ21090	Abq21090 Oligonuc
	24	507.8	35.4	707	6	ABQ21091	Abq21091 Oligonuc
	25	497.8	34.7	1425	4	AA554344	AA554344 Pseudom
	26	497.8	34.7	1425	4	ACA42715	ACA42715 Prokaryot
	27	494.2	34.5	1581	11	ABD17809	ABD17809 Pseudom
	28	493	34.4	1422	8	ACA5258	ACA5258 Prokaryot
	29	465.6	32.5	1407	8	ACA33795	ACA33795 Prokaryot
	30	425.2	29.7	707	6	ABQ21092	Abq21092 Oligonuc
	31	425.2	29.7	707	6	ABQ21093	Abq21093 Oligonuc
	32	406.2	28.3	2055	5	AA589869	AA589869 DNA encod
	33	397	27.7	555	6	ABQ45445	Abq45445 Oligonuc
	34	397	27.7	555	6	ABQ45444	Abq45444 Oligonuc
	35	391	27.3	1251	11	ABD17700	ABD17700 Pseudom
	36	390.4	27.2	1001	4	AA691438	AA691438 Haemophil
	37	390.4	27.2	1001	6	ABK37818	ABK37818 DNA sequ
	38	388.6	27.1	3641	13	ADP05422	ADP05422 Haemophil
	39	350.2	24.4	963	11	ABD17574	ABD17574 Pseudom
	40	340.8	23.8	555	6	ABQ45443	Abq45443 Oligonuc
	41	340.8	23.8	555	6	ABQ45442	Abq45442 Oligonuc
	42	302.6	21.1	706	6	ABQ20706	Abq20706 Oligonuc
	43	302.6	21.1	706	6	ABQ20707	Abq20707 Oligonuc
	44	301	21.0	700	10	ACF66848	ACF66848 Photorhab
	45	286	19.9	948	8	ACA23565	ACA23565 Prokaryot

#### ALIGNMENTS

RESULT 1	AAA65973	standard; DNA; 1434 BP.
ID	AAA65973	
AC	AAA65973;	
DT	05-OCT-2000	(first entry)
DE	E. coli proliferation associated coding sequence SEQ ID NO:165.	
XX	Escherichia coli; E. coli; proliferation; inhibition; screening;	
KW	antimicrobial; bacterial growth; antiseptic therapy; antibacterial; de.	
OS	Escherichia coli.	
XX		
PN	NC0200044906-A2.	
XX		
PD	03-AUG-2000.	
XX		
PF	27-JAN-2000; 2000MC-US002200.	
XX		
PR	27-JAN-1999; 99US-0117405P.	
PA	(BLIT-) ELITRA PHARM INC.	
XX		
PI	Zyckind J, Ohlsen KL, Trawick J, Forsyth RA, Froelich JM;	
PI	Carr GJ, Yamamoto RT, Xu HH;	
DR	WPI; 2000-514822/46.	
XX	P-PSDB; AAB15968.	
PT	Novel polynucleotides and polypeptides associated with microorganism	
PT	proliferation, used to identify inhibitors of bacterial growth and	
XX	proliferation, for use in antiseptic therapy.	
PS	Claim 8; Page 146-147; 316pp; English.	
CC	AAA65809 to AAA65889 and AAA66058 to AAA66138 represent nucleotide	
CC	sequences derived from Escherichia coli which inhibit E. coli	
CC	proliferation. AAA65890 to AAA66055 and AAB15886 to AAB16040 represent	
CC	nucleotide and protein sequences associated with E. coli proliferation.	
CC	AAA66056 and AAA66057 represent primers used for sequencing E. coli	

CC proliferation inhibiting nucleotide inserts in an example from the  
CC present invention. Methods from the present invention can be used to  
CC identify a proliferation- required gene in a microorganism, by contacting  
CC a microorganism with a proliferation- required gene activity inhibitory  
CC nucleic acid identified in another organism, and determining if  
CC inhibition occurs in the second microorganism. The nucleic acid sequences  
CC identified as being required for bacterial growth and proliferation, can  
CC be used for antisense therapy for killing bacteria

8Q Sequence 1434 BP; 328 A; 352 C; 447 G; 307 T; 0 U; 0 Other;

Query Match	100.0%	Score 1434	DB 3	Length 1434
Best Local Similarity	100.0%	Pred. NO. 0		
Matches 1434	0	Mismatches	0	Gaps 0

Qy	1	ATGAAAGTAA	CGCTGCGAAG	AGTTTGAAC	CTGACGAAG	AGATATG	GTGTGTGGTGAAT	GTGCATG	60
Dp	1	ATGAAAGTAA	CGCTGCGAAG	AGTTTGAAC	CTGACGAAG	AGATATG	GTGTGTGGTGAAT	GTGCATG	60
Qy	61	CTGGATCGTTA	CTGTGTAACG	CCCAAC	CAAGTCGTAT	CTGTGCGGAA	GCACGGATG	CCCGTG	120
Dp	61	CTGGATCGTTA	CTGTGTAACG	CCCAAC	CAAGTCGTAT	CTGTGCGGAA	GCACGGATG	CCCGTG	120
Qy	121	GTTAAAGTAA	TACATCGAAG	AGCTCGGGCG	GGCGGCTAA	CGTGGCCGATG	AATATATC	180	
Dp	121	GTTAAAGTAA	TACATCGAAG	AGCTCGGGCG	GGCGGCTAA	CGTGGCCGATG	AATATATC	180	
Qy	181	GCTTCTCGGTG	CTTAATGACG	CCGTGGTGGG	TTGAGGGCA	TTGACATG	CAGACGGCG	240	
Dp	181	GCTTCTCGGTG	CTTAATGACG	CCGTGGTGGG	TTGAGGGCA	TTGACATG	CAGACGGCG	240	
Qy	241	GGCGTAGTAA	TCTGTGGCCG	ACGTCAAC	GTCAATCGA	CTTGTTCTGTAC	CGACG	300	
Dp	241	GGCGTAGTAA	TCTGTGGCCG	ACGTCAAC	GTCAATCGA	CTTGTTCTGTAC	CGACG	300	
Qy	301	CATCCGACAT	TACCAA	TTTACGGGTACT	TTTCCGCAAC	CAACAGCTGAT	CCGTCTG	360	
Dp	301	CATCCGACAT	TACCAA	TTTACGGGTACT	TTTCCGCAAC	CAACAGCTGAT	CCGTCTG	360	
Qy	361	TTTGAAGAAG	GTTCGAGGTG	TGATCCGACG	CCGCTGAC	CGAGCCGATTA	ATCAGGGG	420	
Dp	361	TTTGAAGAAG	GTTCGAGGTG	TGATCCGACG	CCGCTGAC	CGAGCCGATTA	ATCAGGGG	420	
Qy	421	CTGAGTTCAT	TTGGCCGCTGTG	CTTCTGTG	CTAC	TGCGCAAAAGGTG	CGCTGCAAGCGTA	480	
Dp	421	CTGAGTTCAT	TTGGCCGCTGTG	CTTCTGTG	CTAC	TGCGCAAAAGGTG	CGCTGCAAGCGTA	480	
Qy	481	CAGCAGATGAT	CCACTGCGCGGTAA	AGCGGGTGT	CCGGTGTGAT	TGATCCAAAAGT	540		
Dp	481	CAGCAGATGAT	CCACTGCGCGGTAA	AGCGGGTGT	CCGGTGTGAT	TGATCCAAAAGT	540		
Qy	541	ACGGA	TTTTGAGGCTAC	CGCGGGCGCTAC	GCCTTAA	CGCCGATCTCTG	GGAA	600	
Dp	541	ACGGA	TTTTGAGGCTAC	CGCGGGCGCTAC	GCCTTAA	CGCCGATCTCTG	GGAA	600	
Qy	601	GCTGTGT	CGGTAAATGTA	AGACCGAAG	AGATTTGT	TGAGCGCGCATG	AAATGAT	660	
Dp	601	GCTGTGT	CGGTAAATGTA	AGACCGAAG	AGATTTGT	TGAGCGCGCATG	AAATGAT	660	
Qy	661	GCCGATTA	CGAATCTTCG	GCTGTGTGTA	GCCTTCCGAA	CAGGGTATGT	CGCTG	720	
Dp	661	GCCGATTA	CGAATCTTCG	GCTGTGTGTA	GCCTTCCGAA	CAGGGTATGT	CGCTG	720	
Qy	721	CAACCGGGTAA	ACGGCGGCTG	ATAGCCAA	CCCAAGGCG	AGGAAGTAT	ATGACGTTAC	780	
Dp	721	CAACCGGGTAA	ACGGCGGCTG	ATAGCCAA	CCCAAGGCG	AGGAAGTAT	ATGACGTTAC	780	
Qy	781	GGTGGCGGCA	CAACGGTGA	TTGGCGTCT	TGCGCGCA	ACGTGCAAGCGGGTAA	TTCCGCTG	840	
Dp	781	GGTGGCGGCA	CAACGGTGA	TTGGCGTCT	TGCGCGCA	ACGTGCAAGCGGGTAA	TTCCGCTG	840	
Qy	841	GAAAGAC	CTGCTTTT	TGCAATGCGCGCT	GGCGTGGTGT	CGCAACTGGAA	CC	900	
Dp	841	GAAAGAC	CTGCTTTT	TGCAATGCGCGCT	GGCGTGGTGT	CGCAACTGGAA	CC	900	

Db	841	GAAGAAAGCGTCTTCTTTGGCCAAATGCGGCGGGCTGCGGTGGTGGTCCGCAAACTGGGAACC	900
QY	901	TCGACGGTTTCGCCGATCGAGCTGGAAAAATGCTGTACGTGAGACGTGACATACAGGCTTT	960
Db	901	TCGACGGTTTCGCCGATCGAGCTGGAAAAATGCTGTACGTGAGACGTGACATACAGGCTTT	960
QY	961	GGCGTGATGACCCGAAAGAGAACTGAAACGTGCGCGGTAGCGGCAACGCGCTAAACGTGGTGA	1020
Db	961	GGCGTGATGACCCGAAAGAGAACTGAAACGTGCGCGGTAGCGGCAACGCGCTAAACGTGGTGA	1020
QY	1021	AAAGTGTGTGATGACCAACGCGTGTCTTTGACATCTCTGACACCGCGGACAGTCTCTTATCTG	1080
Db	1021	AAAGTGTGTGATGACCAACGCGTGTCTTTGACATCTCTGACACCGCGGACAGTCTCTTATCTG	1080
QY	1081	GCAATATGCCCGCAAGCTGGGTGACCGGCTTGATTTGTCGCGTCAACAGCGATGCTCCACC	1140
Db	1081	GCAATATGCCCGCAAGCTGGGTGACCGGCTTGATTTGTCGCGTCAACAGCGATGCTCCACC	1140
QY	1141	AAACGGCTGAAAGGGGATTTCCCGCCCGGTAAACCCACTCGAACAGGTATGATTTGTCTG	1200
Db	1141	AAACGGCTGAAAGGGGATTTCCCGCCCGGTAAACCCACTCGAACAGGTATGATTTGTCTG	1200
QY	1201	GGCGCACTGGAAGCGGTGACCTGGGTATGTCGTTTGAAGAAGCACGCGCGACGGCTTG	1260
Db	1201	GGCGCACTGGAAGCGGTGACCTGGGTATGTCGTTTGAAGAAGCACGCGCGACGGCTTG	1260
QY	1261	ATCCGCGGGGATCTTTCGCAGATCTGCTGTGTAAAGCGCGCACTTATAACGAGAAGATT	1320
Db	1261	ATCCGCGGGGATCTTTCGCAGATCTGCTGTGTAAAGCGCGCACTTATAACGAGAAGATT	1320
QY	1321	GCCGGGAGTAAAGAAAGTCTGGGCGCAACGGTGGCGAAGTGTGGTGCTCAACTTTGAAGAC	1380
Db	1321	GCCGGGAGTAAAGAAAGTCTGGGCGCAACGGTGGCGAAGTGTGGTGCTCAACTTTGAAGAC	1380
QY	1381	GGTTCCTCGACGACCAATCATCTCAAGAAATCAACAGATTAATAAAGGCTAA	1434
Db	1381	GGTTCCTCGACGACCAATCATCTCAAGAAATCAACAGATTAATAAAGGCTAA	1434

## RESULT 2

ID	AA552567	standard; DNA; 1434 BP

AC AAS52567;

DT 13-FEB-2002 (first entry)

DE E. coli DNA for cellular proliferation protein #289.

KW Antisense; ds; prokaryotic cellular proliferation gene; antibiotic;

XX

**OS Escherichia coli.**

PN W0200170955-A2.

PD 27-SEP-2001

PF 21-MAR-2001; 2001WO-US009180

PR 21-MAR-2000; 2000US-0191078P

PR 26-MAY-2000; 2000US-0207727P

PR 27-NOV-2000; 2000US-0253625P

PR 16-FEB-2001; 2001US-0269308P

PA (ELIT-) ELITRA PHARM INC.

PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GT;

XX







ID ACA18560 standard; DNA; 1434 BP.  
 AC ACA18560;  
 DT 19-JUN-2003 (first entry)  
 DE Prokaryotic essential gene #217.  
 XX  
 KM Antisense; ds; prokaryotic essential gene; cell proliferation;  
 KM drug design; gene.  
 OS Escherichia coli.  
 XX  
 PN W030027183-A2.  
 PD 03-OCT-2002.  
 XX  
 PF 21-MAR-2002; 2002WO-US009107.  
 XX  
 PR 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362639P.  
 XX  
 PA (BLIT-) BLITRA PHARM INC.  
 XX  
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW,  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX WPI; 2003-029926/02.  
 DR P-PSDB; ABU14690.  
 XX  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 PS Claim 14; SEQ ID NO 6430; 1766bp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 643 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) providing a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pat\_sequences  
 XX  
 SQ Sequence 1434 BP; 328 A; 352 C; 447 G; 307 T; 0 U; 0 Other;

Query Match 100.0%; Score 1434; DB 8; Length 1434;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGAAAGTAA	CGCTGCCAGAGTTTGA	CGTGCAGAGTGA	TGTTGGTGTGATG	TGATG	60
DB	1	ATGAAAGTAA	CGCTGCCAGAGTTTGA	CGTGCAGAGTGA	TGTTGGTGTGATG	TGATG	60
QY	61	CTGATGCTTACT	GTGTCACGCCCCCA	CGATGCTATCT	TGCGCGGAAGCG	CGGTCCTG	120
DB	61	CTGATGCTTACT	GTGTCACGCCCCCA	CGATGCTATCT	TGCGCGGAAGCG	CGGTCCTG	120
QY	121	GTTAAAGTAA	TACATCGAAGACGT	CGGCGGCGCG	CGGCTACAGTGCATG	ATATC	180
DB	121	GTTAAAGTAA	TACATCGAAGACGT	CGGCGGCGCG	CGGCTACAGTGCATG	ATATC	180
QY	181	GCTTCTCTCG	GTGCTAAATGCA	CGCTGTGAGGTTGA	CGGCGCATG	CAATGACGCGC	240
DB	181	GCTTCTCTCG	GTGCTAAATGCA	CGCTGTGAGGTTGA	CGGCGCATG	CAATGACGCGC	240
QY	241	GCGCTGAGTAA	TCTCTGCGCGCA	CGTCAAGTCA	ATGCGATCTTCTG	TACCGACG	300
DB	241	GCGCTGAGTAA	TCTCTGCGCGCA	CGTCAAGTCA	ATGCGATCTTCTG	TACCGACG	300
QY	301	CATCCGACATTA	CCAAATTACG	GGGTACTTCCG	CGAACACAG	CGATCGGCTG	360
DB	301	CATCCGACATTA	CCAAATTACG	GGGTACTTCCG	CGAACACAG	CGATCGGCTG	360
QY	361	TTTGAAGAGT	TTTCGAGGTGTTGA	TCCGACCGCTG	GCACGAGGATTA	ATCAGCGC	420
DB	361	TTTGAAGAGT	TTTCGAGGTGTTGA	TCCGACCGCTG	GCACGAGGATTA	ATCAGCGC	420
QY	421	CTGAGTTTCA	TGCGCGCGCTG	GTCTTCTG	ATCAGCCCAAAGT	CGCTGCAAGCTA	480
DB	421	CTGAGTTTCA	TGCGCGCGCTG	GTCTTCTG	ATCAGCCCAAAGT	CGCTGCAAGCTA	480
QY	481	CAGCAGATGA	TCCAACTGCGCGGTAA	AGCGGATTC	CCGGTGTGATG	TCCAAAAGT	540
DB	481	CAGCAGATGA	TCCAACTGCGCGGTAA	AGCGGATTC	CCGGTGTGATG	TCCAAAAGT	540
QY	541	ACCGATTTTGA	AGCGCTACCGCGCGCTA	CGCTGTAA	CGCGCAATCTT	CGGAATTTGA	600
DB	541	ACCGATTTTGA	AGCGCTACCGCGCGCTA	CGCTGTAA	CGCGCAATCTT	CGGAATTTGA	600
QY	601	GCTGTTTGTG	GTAAATGTAAG	ACCGAAGAA	GATTTGTA	AGCGCGCATG	660
DB	601	GCTGTTTGTG	GTAAATGTAAG	ACCGAAGAA	GATTTGTA	AGCGCGCATG	660
QY	661	GCGGATTTAG	CACTCGGCTGTTA	GTAGTACCG	GTTCGAA	CAGGATATG	720
DB	661	GCGGATTTAG	CACTCGGCTGTTA	GTAGTACCG	GTTCGAA	CAGGATATG	720
QY	721	CAACCGGTTAA	CGCGCGCTG	CAATATGCA	ACCAAGCGG	AGATGTA	780
DB	721	CAACCGGTTAA	CGCGCGCTG	CAATATGCA	ACCAAGCGG	AGATGTA	780
QY	781	GCTGCGGCGCA	CACTGATTTG	CGCTGCGGCA	CACTGCA	CGGCTGTA	840
DB	781	GCTGCGGCGCA	CACTGATTTG	CGCTGCGGCA	CACTGCA	CGGCTGTA	840
QY	841	GAGAAAGCCG	CTCTTTGCAATG	CGGCGGCTG	GGTGTGCG	CAAACTGGG	900
DB	841	GAGAAAGCCG	CTCTTTGCAATG	CGGCGGCTG	GGTGTGCG	CAAACTGGG	900
QY	901	TCCACGTTTGC	CGATTCAGTGA	AAATGCTG	TACGTGCA	GATTA	960
DB	901	TCCACGTTTGC	CGATTCAGTGA	AAATGCTG	TACGTGCA	GATTA	960
QY	961	GCGGTATGAC	CGAAGGAATG	AAAGTGTG	CGCGGCGG	GTAAAGTGTG	1020
DB	961	GCGGTATGAC	CGAAGGAATG	AAAGTGTG	CGCGGCGG	GTAAAGTGTG	1020
QY	1021	AAAGTGTATG	CAACGCGTGT	TTTGAATCT	CTGCA	CGCGGCGG	1080

Db 1021 AAGTGGATGACCAACGAGTCTTGGACATCTCGACCGCCGACAGTCTTATCTG 1080  
Qy 1081 GCATATGCCCCGACGCTGGGAGACCGCTTGTGTTGGCCGTCACACAGCATGCTCCACC 1140  
Db 1081 GCATATGCCCCGACGCTGGGAGACCGCTTGTGTTGGCCGTCACACAGCATGCTCCACC 1140  
Qy 1141 AAGCGCTGAAAGGGGATTCGCCGCCGTTAAACCACTCGAACAGCGTATGATGCTG 1200  
Db 1141 AAGCGCTGAAAGGGGATTCGCCGCCGTTAAACCACTCGAACAGCGTATGATGCTG 1200  
Qy 1201 GCGCAGCTGAAAGCGTCTGAGTGTGTTGAAAGAGACACGCGCAGCGCTTG 1260  
Db 1201 GCGCAGCTGAAAGCGTCTGAGTGTGTTGAAAGAGACACGCGCAGCGCTTG 1260  
Qy 1261 ATCCCGCGGATCTTCCAGATCTGCTGTGAAAGCGCGCCGCTTAAACCGAAGAGATT 1320  
Db 1261 ATCCCGCGGATCTTCCAGATCTGCTGTGAAAGCGCGCCGCTTAAACCGAAGAGATT 1320  
Qy 1321 GCCCGAGTAAAGAGTCTGGGCGCAACGCTGCGAGTGTGCTCAACTTGAAGAC 1380  
Db 1321 GCCCGAGTAAAGAGTCTGGGCGCAACGCTGCGAGTGTGCTCAACTTGAAGAC 1380  
Qy 1381 GGTGCTCGACGACCAATCATCAAGAGATCCACAGATAAAGGCTAA 1434  
Db 1381 GGTGCTCGACGACCAATCATCAAGAGATCCACAGATAAAGGCTAA 1434

RESULT 5  
ADH88994  
ID ADH88994 standard; DNA; 1434 BP.

AC ADH88994;

DT 22-Apr-2004 (first entry)

XX Escherichia coli K1 S26 mutant coding sequence, SEQ ID 28.

XX Pathogenic microorganism; K1 S26; mutant; pathogenicity; virulence;  
KW bacterial infection; extra-intestinal infection; enterobacterium; gene;  
db.

OS Escherichia coli.  
OS Synthetic.

XX Key location/Qualifiers  
FT CDS 1..1434  
FT /\*tag= a  
FT /product= "K1 S26 mutant"

XX MO2004005535-A2.

XX 15-JAN-2004.

XX 09-JUL-2003; 2003WO-EP008209.

XX 09-JUL-2002; 2002FR-0008636.

XX (MUTA-) MUTABILIS SA.

XX Becalch S;

XX MPI; 2004-091381/09.

XX P-PSDB; ADH88994.

XX Identifying and selecting a gene required for the proliferation in vivo  
PT of a pathogenic microorganism comprises determining the virulence of  
PT mutant genes on an experimental model of infection, and their effect on  
PT enteric colonization.

XX Claim 5; SEQ ID NO 28; 60BP; English.

XX The present invention relates to a method for identifying and selecting a

CC gene required for the proliferation in vivo of a pathogenic  
CC microorganism. The method comprises determining the virulence of mutant  
CC genes on an experimental model of infection, and their effect on enteric  
CC colonization in an axenic mouse model. ADH88994-ADH88991 and ADH88987-  
CC ADH88996 are pathogenicity or virulence targets and their coding  
CC sequences which were used in the method of the invention. Compounds  
CC capable of inhibiting pathogenicity or virulence target expression are  
CC useful for developing medicinal products for inhibiting a bacterial  
CC infection, in particular an extra-intestinal infection in the case of  
CC enterobacteria.

XX Sequence 1434 BP; 328 A; 352 C; 447 G; 307 T; 0 U; 0 Other;

Qy Query Match 100.0%; Score 1434; DB 12; Length 1434;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGAAAGTAAACGCTGCCAGATTTGAACGTGCAGAGTATGCTGTTGATGATG 60  
Db 1 ATGAAAGTAAACGCTGCCAGATTTGAACGTGCAGAGTATGCTGTTGATGATG 60  
Qy 61 CTGATCTTACCTGATACGCGCCCAAGCTGATCTGCGCCGAGGCGCGCGCTG 120  
Db 61 CTGATCTTACCTGATACGCGCCCAAGCTGATCTGCGCCGAGGCGCGCGCTG 120  
Qy 121 GTTAAAGTAAATACATCGAAGACGTCCGCGCGCGCTTAAACGTGCGATGATATC 180  
Db 121 GTTAAAGTAAATACATCGAAGACGTCCGCGCGCGCTTAAACGTGCGATGATATC 180  
Qy 181 GCTTCTCTCGGCTCTAATGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240  
Db 181 GCTTCTCTCGGCTCTAATGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240  
Qy 241 GCGCTGATTAATCTCGCGCGCGAGCTCAACGTCAAAATGCACTTCGTTCTGACGACG 300  
Db 241 GCGCTGATTAATCTCGCGCGCGAGCTCAACGTCAAAATGCACTTCGTTCTGACGACG 300  
Qy 301 CATCCGACATTAACCAATTAACGAGTATCTTCCCGCAACCAACAGCTGATCCGTGAT 360  
Db 301 CATCCGACATTAACCAATTAACGAGTATCTTCCCGCAACCAACAGCTGATCCGTGAT 360  
Qy 361 TTGAAAGAGTTCGAAAGTGTGATCCGACGCGCTGACAGAGGATTAATCAAGCG 420  
Db 361 TTGAAAGAGTTCGAAAGTGTGATCCGACGCGCTGACAGAGGATTAATCAAGCG 420  
Qy 421 CTGAGTTCGATGAGCGCGCTGCTTCTGATCACTGACCAAGGTGCTGCGCAAGCT 480  
Db 421 CTGAGTTCGATGAGCGCGCTGCTTCTGATCACTGACCAAGGTGCTGCGCAAGCT 480  
Qy 481 CAGCAGATGATCCAACTGCGCGCTTAAAGCGGTGTTCCGCTGCTGATTCAAAAGT 540  
Db 481 CAGCAGATGATCCAACTGCGCGCTTAAAGCGGTGTTCCGCTGCTGATTCAAAAGT 540  
Qy 541 ACCGATTTTGAAGCGCTACCGCGCGCTGCTGCTTAAACCGGATTCCTGGAATTTGAA 600  
Db 541 ACCGATTTTGAAGCGCTACCGCGCGCTGCTGCTTAAACCGGATTCCTGGAATTTGAA 600  
Qy 601 GCTGTTGCTGTAATGTAAGACCGAAGAGATTTTGAAGCGCGCATGAATCTGATT 660  
Db 601 GCTGTTGCTGTAATGTAAGACCGAAGAGATTTTGAAGCGCGCATGAATCTGATT 660  
Qy 661 GCCGATTAACAACTCTGCTGCTGTTAGTACCGGTTCCGAACAGGATATGCTGCTG 720  
Db 661 GCCGATTAACAACTCTGCTGCTGTTAGTACCGGTTCCGAACAGGATATGCTGCTG 720  
Qy 721 CAACCGGTTAAAGCGCGCTGCTATATGCCAACCAAGCGGAGATGATACGTTACC 780  
Db 721 CAACCGGTTAAAGCGCGCTGCTATATGCCAACCAAGCGGAGATGATACGTTACC 780  
Qy 781 GGTGCGGCGGACACGAGTATGCGCTGCTGCGCGCAACGCTGACAGCGGATTAATGCTG 840  
Db 781 GGTGCGGCGGACACGAGTATGCGCTGCTGCGCGCAACGCTGACAGCGGATTAATGCTG 840



```
Dh 361 TTTGAGAGAGATTGAGAGCGTTGATCCGAGCGCTGACGAGCGTATCAACAGGCG 420
Qy 421 CTGAGTTGATTTGGCGGCTGGTCTTCTGACTACGCAAGAGGCGCTGGCAAGGTA 480
Dh 421 CTGGGCAATATTTGGCGGCTGGTCTTCTGACTACGCAAGAGGCGCTGGCAAGGTA 480
Qy 481 CAGGAGATGATCACTGGCGCTGTAAGCGGGGTTCGCGTGTGATGATCCAAAAGT 540
Dh 481 AAGAGATGATTTCAAGCTGGCGCTGTAAGCGCAACGCTCCGGTGTGATGATCCAAAAGG 540
Qy 541 ACCGATTTTGAAGCGCTACCGCGGGCTACGCTTTTAAGCGCGAATCTTCGGAATTTGAA 600
Dh 541 ACCGATTTTGAAGCGCTATCGCGGCGCAACCTGCTGACGCGCAACCTCTCCGAGTTTGA 600
Qy 601 GCTGTGTGCGTAAATGTAAGACGGAAGAGAATGTTGAGCGCGGATGAACTGATT 660
Dh 601 GCTGTGTGCGGCAAGTGCAGAACCGAGAGAGAGCTGGTTGAGCGCGGATGAACTGATT 660
Qy 661 GCCGATTAAGAACTCTCGGCTCTGTAGTGAACCGTTCCGAAAGGATATGCTGCTG 720
Dh 661 GCCGATTTGAGCTGTCCGCTGTGCTGACCCGCTCCAGCAGGGGATGACGCTGCTG 720
Qy 721 CAACCGGGTAAAGCGCGCTGCTGATATGCAACCAAGCGCGAAGTGTATGAGTTACC 780
Dh 721 CAGCGGGCAAAAGCGCGCTGCTGATATGCAACCAAGCGCGAAGTGTATGAGTTACC 780
Qy 781 GGTGCGGGCGACACGAGTATGAGCGTCTGCGGCAACGCTGAGCGAGGATATTCGCTG 840
Dh 781 GGTGCGGGTATGAGTATGAGCGTCTGCGGCAACGCTGAGCGAGGATATTCGCTG 840
Qy 841 GAAGAGCGCTGCTTCTTTGCCAATGCGGCGGCTGCGTGTGCTGCGCAACCTGAGAAC 900
Dh 841 GAAGAGGCGCTACTTCTGCTAACGCGCGGCGGCTGCGTGTGCTGCGTAACTGAGTACC 900
Qy 901 TCCACGCTTTGCGCGATGAGACTGGAATAATCTGTAGCTGAGACGTGCAGATACAGGCTTT 960
Dh 901 TCAACCGCTTTGCGCAATGAGACTGGAATAACCGGTGTGCGGCTGCGGATACAGGCTTTT 960
Qy 961 GAGCGTATGACCGAAGAGAACTGAAAGCTGCGCGGAGCGGAGCGCGTAAAGCTGTGAA 1020
Dh 961 GAGCGTATGACCGAAGAGAACTGAAAGTGGCTGCGCGGAGCGGAGCGCGTGTGAA 1020
Qy 1021 AAAGTGTGATGACCAACGAGTGTCTTTGACATCTTCGACGCGCGGACAGTCTTTATCTG 1080
Dh 1021 AAAGTGTGATGACCAACGAGGCTGTTGACATCTTCGACGCGCGGACAGTCTTTATCTG 1080
Qy 1081 GCAAAATGCCCGCAAGCTGGGAGACCGCTTGAATTTGTCGCTGCAACAGGATGCTCCACC 1140
Dh 1081 GCGAAACGCGCGCAAGCTGGGAGATCGCTGATTTGAGGCTTAAACAGGATGATCGACG 1140
Qy 1141 AAAGCGCTGAAGAGGGGATTTCCCGCCCGGTAAACCACTCGAAACAGCGTATGATTTGCTG 1200
Dh 1141 AAGCGCTGAAGAGGGGAAACCGCTGCGGTGAACCGCTCGAGCAGCGATATTTGATCTC 1200
Qy 1201 GCGCGACTGAAGCGGCTGAGCTGGTGTGTTGAAAGAGACACGCGCGACGCTTG 1260
Dh 1201 GGTGCGCTGAAGCGGCTGAGCTGGTGTGTTGAAAGAGATACCCCGCAGCGCTG 1260
Qy 1261 ATCCCGCGGATCTTGGCAGATCTGCTGTGAAAGCGCGCGACTATTAACCAAGAAAGATT 1320
Dh 1261 ATTCGCGGCAATTTGCTGCGAGCTGTGTGTAAGAGGTGGGATTTCAAGCGCGAGCAATTC 1320
Qy 1321 GCGCGAGTAAAGAGTCTGGGCGCAACGCTGGCGAGAGTGTGTGCTCAACTTTGAAGAC 1380
Dh 1321 GCGCGAGAGAGAGTCTGGGCGCAACGCGCGAGAGTGTGTGCTCAACTTTGAAGAC 1380
Qy 1381 GGTGTCTGACGACCAATCATCAAGAAAGATCCAAACGAGATA 1423
Dh 1381 GGTGTCTCAACCAACATCATCAAGAAAGATCCAAAGAGACA 1423
```

RESULT 7  
ACH96242

```
ID ACH96242 standard; DNA; 1500 BP.
XX
AC ACH96242:
XX
DT 29-JUL-2004 (first entry)
XX
DE Klebsiella pneumoniae polynucleotide seqid 2037.
XX
KW Recombinant expression vector; transcription regulatory element;
XX Klebsiella pneumoniae protein; antibacterial; Vaccine; gene; de.
XX
OS Klebsiella pneumoniae.
XX
PN US6610836-B1.
XX
PD 26-AUG-2003.
XX
PE 27-JAN-2000; 2000US-00489039.
XX
PR 29-JAN-1999; 99US-0117747P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Breton GU, Osborne M;
XX
DR WPI; 2003-895346/82.
XX
PT P-PSDB; AB062691.
XX
PT New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for
XX preparing a vaccine composition against Klebsiella pneumoniae.
XX
PS Disclosure; SEQ ID NO 2037; 932pp; English.
XX
CC The invention describes a new isolated nucleic acid encoding a Klebsiella
XX pneumoniae polypeptide. Also described are: a recombinant expression
XX vector comprising the nucleic acid, operably linked to a transcription
XX regulatory element; and a cell comprising the recombinant expression
XX vector. The nucleic acid is useful for preparing a vaccine composition
XX against Klebsiella pneumoniae. This sequence encodes a Klebsiella
XX pneumoniae polypeptide of the invention
XX
SQ Sequence 1500 BP; 308 A; 416 C; 497 G; 279 T; 0 U; 0 Other;
XX
Query Match 70.6%; Score 1011.8; DB 11; Length 1500;
Best Local Similarity 81.9%; Pred. No. 7e-290;
Matches 1166; Conservative 0; Mismatches 257; Indels 0; Gaps 0;
Qy 1 ATGAAAGTAAAGCTGCCAGAGTTTGAACGTGACGAGATGATGTTGGTATGATG 60
Dh 67 ATGAAAGTAAAGCTGCCAGAGTTTGAACGTGACGAGATGATGTTGGTATGATG 126
Qy 61 CTGAGTCTTACGTGTAAGCGGCGCCACCAAGTGTATCTGCGCGGAAGCGCGTGCCTG 120
Dh 127 CTGAGTCTTACGTGTAAGCGGCGCCACCAAGTGTATTTCCCGAAGCGCGTGCCTG 186
Qy 121 GTTAAAGTAAATCAATCGAAGACGTCCGCGGCGCGCTAAAGTGGAGTAATATC 180
Dh 187 GTGAGGTGGAATAATTCGAAGACGTCTGCGCGCGCGCAACGTAGGAGTGAACATT 246
Qy 181 GCTTCTCTCGGTGTATGACAGCTGCTGGTGGTGTGACGGGCAATTAAGATGACGCGC 240
Dh 247 GCTCTCGTGGGGCAACTTCGCGCTGTGTGGAATTGACCGGAAATGATGACGCGCGC 306
Qy 241 GCGCTAGTAATCTCTGCGCGACGTCAAGTCAAAATGCGACTTCTGTTTACGAGC 300
Dh 307 GCGCTAGGCGAGCGCTGCGCAACGTCAATGTAAAGTGCACATTTCTCTCGCTCCGACT 366
Qy 301 CATCCGACATTAACCAATTAACGAGTACTTTCCGCAACCAAGACGTATCGTCTGAT 360
Dh 367 CACCGGATATCAACAAAGCTGGGGTGTGTGCGGAAATCAACAGCTGATCCGCTGAC 426
Qy 361 TTTGAAAGAGTTTCAAGAGTGTGATCCGACGCGCTGACAGCGGATTAATCAAGCG 420
```

```

Db 427 TTGAAAGAGGCTTCTCCGGGCTGATCCGACGGCGATGACGAGCGCATTCAGACGGCG 486
Qy 421 CTGAGTTGATTTGGCGGCTGCTGCTTTCTGACTACGCAAAAGTGGCTGCAAGGTA 480
Db 487 CTGGGCTCTATTTGGCGGCTGCTGCTTTCTGACTACGCAAAAGGCGCTGACCAAGGCTC 546
Qy 481 CAGCAGATGATTCAGACTGCGCGCTGTAAGCGGGTGTTCGGGCTGATTTGATTCAAAAGT 540
Db 547 CAGACATGATTCAGACTGCGCGCTGTAAGCGGGTGTTCGGGCTGATTCAAAAGGCTC 606
Qy 541 ACCGATTTGAGCGCTACCGCGGCGCTACCGCTGTTAAACCGAAATCTCTCGAAATTTGAA 600
Db 607 ACCGATTTGAGCGCTACCGCGGCGCTACCGCTGTTAAACCGAAATCTCTCGAAATTTGAA 666
Qy 601 GCTGTTGCTGTAATGTAAGACCGAAAGAGATTTGAGCGGCTGATTTGAAATCTGATT 660
Db 667 GCGGTGTTGAGCGCTACCGCGGCGCTACCGCTGTTAAACCGAAATCTCTCGAAATTTGAA 726
Qy 661 GCCGATTTACGAACTCTCGGCTCTGTTAGTACCGGCTTCCGAAACAGGGTATGCTGCTG 720
Db 727 GCCGATTTACGAACTCTCGGCTCTGTTAGTACCGGCTTCCGAAACAGGGTATGCTGCTG 786
Qy 721 CAACCGGCTAAAGCGCGCTGCAATATGCAACCCAGCGCAGAAATGTAAGCTTACC 780
Db 787 CAGCGGGGACGTCGCGGCTGCAATATGCGCAACCCAGCGCAGAAATGTAAGCTTACC 846
Qy 781 GGTGCGGCGCAACAGGCTGATTTGGCTCTGCGCGCAACGCTGCGACGGGTAATTCGCTG 840
Db 847 GCGCGCGCGCAACAGGCTGATTTGGCTCTGCGCGCAACGCTGCGACGGGTAATTCGCTG 906
Qy 841 GAAGAACCTGCTCTTTGCAATGCGGCGCTGCGGCTGATGCTGCGCAACCTGGAAC 900
Db 907 GAAGAACCTGCTCTTTGCAATGCGGCGCTGCGGCTGATGCTGCGCAACCTGGAAC 966
Qy 901 TCACCGGCTTTCGCGGCTGATGCTGCAAAATGCTGATGCTGCAAGTACAGGCTTT 960
Db 967 TCACCGGCTTTCGCGGCTGATGCTGCAAAATGCTGATGCTGCAAGTACAGGCTTT 1026
Qy 961 GCGCTGATGACCGAAAGAGAACTGAACTGCGCTGACGGGCAAGGCTTAAACGCTGGA 1020
Db 1027 GCGCTGATGACCGAAAGAGAACTGAACTGCGCTGACGGGCAAGGCTTAAACGCTGGA 1086
Qy 1021 AAATGATGATGACCAACGCTGCTTTGACATCTCTGACGCGCGGCAAGCTCTTTATCTG 1080
Db 1087 AAATGATGATGACCAACGCTGCTTTGACATCTCTGACGCGCGGCAAGCTCTTTATCTG 1146
Qy 1081 GCAAAATGCGCGCAAGCTGCTGATGCTGCAAAATGCTGATGCTGCAAGCTCTTCAAC 1140
Db 1147 GCGAATGCGCGCAAGCTGCTGATGCTGCAAAATGCTGATGCTGCAAGCTCTTCAAC 1206
Qy 1141 AAACGCTGAAAGGGGATTTCCCGCCCGGTAACCCCACTCGAACAGGCTATGATTTGCTG 1200
Db 1207 AAACGCTGAAAGGGGATTTCCCGCCCGGTAACCCCGCTGGAACCGCGATGATGCTGCTG 1266
Qy 1201 GGGCGCATGGAAGGGCTGACCTGGTATGCTGTTGAAAGGACAGCGCGCAGCGCTTG 1260
Db 1267 GGGCGCATGGAAGGGCTGACCTGGTATGCTGTTGAAAGGATCTCGCAGCGCGCTG 1326
Qy 1261 ATGCGCGGATCTTTCAGATCTGCTGATGAAAGCGCGCACTATTAACAGAGAGATT 1320
Db 1327 ATGCGCGGATCTTTCAGATCTGCTGATGAAAGCGCGCACTATTAACAGAGAGATT 1386
Qy 1321 GCGCGGATTAAGAAAGTCTGGGCGCAAGCTGCGAAGTGTGGTCTCAACTTTGAAGAC 1380
Db 1387 GCGCGGATTAAGAAAGTCTGGGCGCAAGCTGCGAAGTGTGGTCTCAACTTTGAAGAT 1446
Qy 1381 GGTGCTGACGACCAATCATCAAGAAAGATTCACAGAGATA 1423
Db 1447 GGTGCTGACCAATCATCAAGAAAGATTCACAGAGATA 1489

```

RESULT 8  
AAS56024

```

ID AAS56024 standard; DNA; 1434 BP.
XX
AC AAS56024;
XX
DT 13-FEB-2002 (first entry)
XX
DE Salmonella typhi DNA for cellular proliferation protein #57.
XX
KW Antisense; ds; prokaryotic cellular proliferation gene; antibiotic;
XX antibiotic; drug design.
XX
OS Salmonella typhi.
XX
PN M02001.70955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001MO-US009180.
XX
PR 21-MAR-2000; 2000US-0191078P.
PR 23-MAY-2000; 2000US-0206848P.
PR 26-MAY-2000; 2000US-0207727P.
PR 23-OCT-2000; 2000US-0242578P.
PR 27-NOV-2000; 2000US-0253625P.
PR 22-DEC-2000; 2000US-0257931P.
PR 16-FEB-2001; 2001US-0269308P.
XX
PA (BLIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zykkind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
XX
DR P-PSDB; AAU38165.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids.
XX
XX
PS Claim 27; SEQ ID NO 9661; 511bp; English.
XX
XS
XS The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the genes,
CC their use in the discovery of novel antibiotics, the essential genes,
CC themselves and the encoded proteins. The prokaryotes used are Escherichia
CC coli, Streptococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
CC useful for the identification of potential new targets for antibiotic
CC development. The antisense nucleic acids can also be used to identify
CC proteins used in proliferation, to express these proteins, and to obtain
CC antibodies capable of binding to the expressed proteins. The proteins can
CC be used to screen compounds in rational drug discovery programmes. The
CC antisense nucleic acid sequence is also useful to screen for homologous
CC nucleic acids which are required for cell proliferation in a wide variety
CC of organisms. The present sequence encodes an essential prokaryotic
CC cellular proliferation protein. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pat_sequences
XX
SQ Sequence 1434 BP; 311 A; 360 C; 468 G; 295 T; 0 U; 0 Other;

```

Query Match 70.3%; Score 1008.6; DB 4; Length 1434;  
Best Local Similarity 81.8%; Pred. No. 6.1e-289;  
Matches 1164; Conservative 0; Mismatches 259; Indels 0; Gaps 0;

```

Qy 1 ATGAAGTAAAGCTGCGACAGATTGAACTGACGAGATGCTGCTGCTGATGATGATG 60
Db 1 ATGAAGTAAATGTCAGAGTTTGAACGTGACGGCTATGCTGCTGCTGATGATGATG 60
Qy 61 CTGATGCTTACTGTTACGCGCCCAACAGTCTGATCTCGCGGAGCGCGGCTGCGCG 120
Db 61 CTGATGCTTATGTTATGCGCCCACTGCGCGGATTTACCGGAGCGCGGCTGCGCG 120

```



```

Qy 121 GTTAAAGTAATACATCGAAGAAAGTCCGGGCGCGCGGTAAAGTGGCGAATATATC 180
Db 121 GTTAAAGTAATACATCGTGAAGAAAGTCCGGGCGCGCGGTAAAGTGGCGAATATATC 180
Qy 181 GCTTCTCGGTGCTAATGACGCGCTGGTGGGTGACGGGCAATTAAGATGACGGCGC 240
Db 181 GCGTCTCGGGAGGAAAGCGCGCTGGTGGGTGACGGGCAATTAAGATGACGGCGCGC 240
Qy 241 GCGGTAGTAATATCTGGGCGAAGTCAAGTCAATGCGCACTTGTTTGTATCCGACG 300
Db 241 GCGGTAGCAAAAGCTGGCGAAGTCAATGCAATGCGCACTTGTTTGTATCCGACG 300
Qy 301 CATCGAAGCAATTAACAAATTAACGGGTACTTTCGCGCAACCAACAGTATCCGTGAT 360
Db 301 CATCGAAGCAATTAACAAATTAACGGGTACTTTCAGGTATCAAGAGTCAATTCGTGAT 360
Qy 361 TTGAAAGAGGTTTCGAAGGTGTTGATCCGACCGCGCTGCAAGACCGAATTAATCAGCG 420
Db 361 TTTGAAAGAGGCTTTGAGGGGTTGACCCGACCGCGTTCATGACGCTATCAACAGCG 420
Qy 421 CTGAGTTGATTTGGCGCGCTGGTGTCTGACTACGCAAGGAGGCGCTGGCAAGGCTA 480
Db 421 CTGGGATTCATCGCGCGCTGGTGTCTGACTATTCGCAATTAAGCAAGGCGCTGACAGCG 480
Qy 481 CAGCAGATGATCAACTGGCGCGTAAAGCGGCTGTTCCGCTGCTGATTTGATCCAAAGGT 540
Db 481 CAGACTATGATTTCCCTAGCGCGCAAGCGGCGGCTGCTCATCGATCCGAAAGGA 540
Qy 541 ACCGATTTTGAAGCGCTACCGCGCGCTACGCTGTTAACGCGCAATCTCTCGGAATTTGAA 600
Db 541 ACGGATTTTGAAGCTTTACCGCGCGCTACGCTGCTGACGCAAACTTTCTGAATTTGAG 600
Qy 601 GCTGTTGCTGTAATATGTAAGACCAAGAAAGATTTGTTAGCGCGGCAATGAATCTGAT 660
Db 601 GCGGTTGCGGGAAATGTAAGCAAGAAAGATTTGTTAGCCGCGATTAACCTCAAT 660
Qy 661 GCCGATTAAGAACTCTCGGCTCTGTTAGTGAACCGCTTCGAAACGGGTATGCTGCTG 720
Db 661 GCCGATTAAGAACTCTCGGCTCTGTTAGTGAACCGCTTCGAAACGGGTATGCTGCTG 720
Qy 721 CAACCGGTTAAAGCGCGCTGATATGCAACCGCAAGCGGCAAGATGATGAAGTTAAC 780
Db 721 CAACCGAATTAAGCGCGCTGATATGCAACCGCAAGCGGCAAGATGATGAAGTTAAC 780
Qy 781 GGTGCGGCGCAACAGCGTATGCGCTCTGCGCGCAACGCTGCGAATTTCCGCTG 840
Db 781 GGTGCGGCGCAATCGATGATGCGCTCTGCGCGCAACGCTGCGGCGGAAATACCTG 840
Qy 841 GAAGAAAGCTCTCTTTCGCAATGCGCGCTGCGCTGCTGCTGCGCAATCTGCGAACC 900
Db 841 GAGGAGCGGTGTTATTTTCGCAATGCGCGCTGCGCTGCTGCTGCTGCGAATCTGCGAACC 900
Qy 901 TCCACGCTTTGCGCGATGCACTGGAATATCTGTAAGTGAAGTGAAGTGAAGTGAAGT 960
Db 901 TCAACGCTTTTCCCTATGAGCTGGAATATCTGTAAGTGAAGTGAAGTGAAGTGAAGT 960
Qy 961 GCGGTGATGCGCAAGCAAGAACTGAAGTGGCGCTGAGCGGCGCGCTGAACGTTGATA 1020
Db 961 GCGGTGATGCGCAAGCAAGAACTGAAGTGGCGCTGAGCGGCGCGCTGAACGTTGATA 1020
Qy 1021 AAAGTGTGATGACCAACGCTGCTTTGACATCTGCAACGCGCGGCAAGTCTTTATCTG 1080
Db 1021 AAAGTGTGATGACCAACGCTGCTTTGACATCTGCAACGCGCGGCAAGTCTTTATCTG 1080
Qy 1081 GCAAAATGCGCGCAAGCTGCTGAGACGCTGATGTTGTCGCTGCAACAGGAGCTTCAACC 1140
Db 1081 GCGAAAGCGCGCAAGCTGCTGAGACGCTGATGTTGTCGCTGCAACAGGAGCTTCAACC 1140
Qy 1141 AAAGCGCTGAAGAGGAGTTCGCGCGCGGTAAACCACTCGAAGACGCTGATTTGCTG 1200
Db 1141 AAAGCTGCAAGAGGAGAGCGCTGCGGTAAATCCGCTCGAAGACGCTGATTTGCTG 1200
Qy 1201 GCGCACTGGAAGCGGCTGCACTGGGTATGCTGTTGAAGAGACACGCGCGACGCTTG 1260

```

```

Db 1201 GCGCGCTGAGATCGGCTCGACTGCTGTTGCTCTTTGAAAGAGATACGCGCAACGACTG 1260
Qy 1261 ATGCGCGGATCTTGGCAATCTGCTGTTGAAGGGGCGGCACTATAACGAAAGAT 1320
Db 1261 ATGCGGATTTCTGCGGATCTGCTGTTGAAGGGGCGGCACTATAAGCGAAGATC 1320
Qy 1321 GCGGAGTGAAGATCTGCGGCAACGCTGCGGCAAGTGTGCTCACTTTGAAGAC 1380
Db 1321 GCGGAGCAAGAAAGTCTGCGGCAACGCGGCGGCAAGTCAATGCTGCAACTTGAAGAT 1380
Qy 1381 GGTGCTCGACGACCAATCATCAAGAGATCAACAGATA 1423
Db 1381 GGTGCTCGACGACCAATCATCAAGAGATCAACAGATA 1423

RESULT 9
ACAS1424
ID ACAS1424 standard; DNA; 1434 BP.
XX
ACAS1424;
AC 19-JUN-2003 (first entry)
DT
XX
DE Prokaryotic essential gene #33081.
XX
KW Antisense; ds; prokaryotic essential gene; cell proliferation;
XX drug design; gene.
XX
OS Salmomella typhl.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (BLIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KU, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
XX WPI; 2003-029926/02.
XX P-PSDB; ABU47554.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 14; SBQ ID NO 39294; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 613 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an

```



organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [http://wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences)

Sequence 1434 BP; 311 A; 360 C; 468 G; 295 T; 0 U; 0 Other;

Query Match 70.3%; Score 1008.6; DB 8; Length 1434;  
Best Local Similarity 81.8%; Pred. No. 6.1e-289;  
Matches 1164; Conservative 0; Mismatches 259; Indels 0; Gaps 0;

QY 1 ATGAAGTAACTGCTGACAGATTGAACTGACAGAGTATGTGTGTGATGATG 60  
DB 1 ATGAAGTAACTGCTGACAGATTGAACTGACAGAGTATGTGTGTGATGATG 60  
QY 61 CTGATGTTTCTGCTGACAGATTGAACTGACAGAGTATGTGTGTGATGATG 120  
DB 61 CTGATGTTTCTGCTGACAGATTGAACTGACAGAGTATGTGTGTGATGATG 120  
QY 121 GTTAAAGTAACTGCTGACAGATTGAACTGACAGAGTATGTGTGTGATGATG 180  
DB 121 GTTAAAGTAACTGCTGACAGATTGAACTGACAGAGTATGTGTGTGATGATG 180  
QY 181 GCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240  
DB 181 GCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240  
QY 241 GCGCTGAGTAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300  
DB 241 GCGCTGAGTAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300  
QY 301 CATCGACATTAACCAATTACGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360  
DB 301 CATCGACATTAACCAATTACGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360  
QY 361 TTGGAAGAGGTTTGAAGGTGTTGATCCGACGCGCTGACGAGGATTAATCAAGCG 420  
DB 361 TTGGAAGAGGTTTGAAGGTGTTGATCCGACGCGCTGACGAGGATTAATCAAGCG 420  
QY 421 CTGAGTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480  
DB 421 CTGAGTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480  
QY 481 CAGAGATGATCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540  
DB 481 CAGAGATGATCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540  
QY 541 ACCGATTTTGAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600  
DB 541 ACCGATTTTGAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600  
QY 601 GCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660  
DB 601 GCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660  
QY 661 GCGGATTAACGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720  
DB 661 GCGGATTAACGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720  
QY 721 CAACCGGATTAACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780  
DB 721 CAACCGGATTAACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780

QY 781 GGTGCGGCGACACGAGGATTTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840  
DB 781 GGTGCGGCGACACGAGGATTTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840  
QY 841 GAAGAACCTGCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900  
DB 841 GAAGAACCTGCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900  
QY 901 TCAACGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960  
DB 901 TCAACGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960  
QY 961 GCGGATTAACGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020  
DB 961 GCGGATTAACGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020  
QY 1021 AAAGTGTGATGACCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080  
DB 1021 AAAGTGTGATGACCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080  
QY 1081 GCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140  
DB 1081 GCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140  
QY 1141 AAACGCTGAAAGGGAATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200  
DB 1141 AAACGCTGAAAGGGAATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200  
QY 1201 GCGGCACTGAAAGGGAATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260  
DB 1201 GCGGCACTGAAAGGGAATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260  
QY 1261 ATGCGGGAATCTTCCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320  
DB 1261 ATGCGGGAATCTTCCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320  
QY 1321 GCGGGAATTAAGGGAATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380  
DB 1321 GCGGGAATTAAGGGAATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380  
QY 1381 GGTGCTGACGACCAATCATCAAGAGATCAAGAGATCAAGAGATCAAGAGATCAAGAGATCAAGAGAT 1423  
DB 1381 GGTGCTGACGACCAATCATCAAGAGATCAAGAGATCAAGAGATCAAGAGATCAAGAGATCAAGAGAT 1423

RESULT 10  
AC36223  
ID AC36223 standard; DNA; 1431 BP.  
XX  
AC36223;  
XX  
AC36223;  
XX  
19-JUN-2003 (first entry)  
XX  
DE Prokaryotic essential gene #17880.  
XX  
KM Antisense; ds; prokaryotic essential gene; cell proliferation;  
XX drug design; gene.  
XX  
OS Klebsiella pneumoniae.  
XX  
PN K0200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002MO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX

PA (BLIT-) ELITRA PHARM INC.  
 XX Mang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JM;  
 PI Mail D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 DR WPI; 2003-029926/02.  
 XX P-PSDB; ABU32353.  
 XX  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 PS  
 XX Claim 14; SEQ ID NO 24093; 1766bp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 CC  
 XX  
 SQ Sequence 1431 BP; 285 A; 407 C; 480 G; 259 T; 0 U; 0 Other;  
 Query Match 70.2%; Score 1007; DB 8; Length 1431;  
 Best Local Similarity 81.7%; Pred. No. 1,8e-288;  
 Matches 1163; Conservative 0; Mismatches 260; Indels 0; Gaps 0;

D	301	CAACCCGACTATCAACAGAGTGGCGGCTGTGTCGCGAATCAACAGCTGAATCCGCTCGAC	360
Q	361	TTTGAAGAGGTTTCGAAGGTGTGATCCGACCGGTGACAGAGCGAATTAATCAGCG	420
D	361	TTTGAAGAGGCTTCTCCGCGCTGATCCGACCGAATGACAGAGCATTCAGCAAGCG	420
Q	421	CTGAGTTGCAATTCGCGCGCTGTGTCTTTCATGACGCGCAAGGTGCGTGCAGAGCTA	480
D	421	CTGGGGCTATATTCGCGCGCTGTGTCTTTCATGACGCGCAAGGTGCGTGCAGAGCTA	480
Q	481	CAGCAGATGATCCAACTGCGCGGTAAAGCGGCTGTTCGGTCTGATTAATCCAAAAGT	540
D	481	CAGCAGATGATCCAACTGCGCGGTAAAGCGGCTGTTCGGTCTGATTAATCCAAAAGT	540
Q	541	ACCGATTTTGAAGGCTTACCGCGCGGTGACGCTGTAAACGCGAATCTCGAATTTGAA	600
D	541	ACCGATTTTGAAGGCTTACCGCGCGGTGACGCTGTAAACGCGAATCTCGAATTTGAA	600
Q	601	GCTGTTGTCCGTAAATGTAAAGACCGAAGAGATTGTAGCGCGCATGAAACTGAT	660
D	601	GCGGTGTGGGCAAGTCCAGATGAAAGCGCAATGTTAGAGCGGCATGAAAGCTATC	660
Q	661	GCGAATTAAGACTCTCGGCTCTGTTAATGACCCGTTCCGAACAGGATATGCTGCTG	720
D	661	GCGAATTAAGACTCTCGGCTCTGTTAATGACCCGTTCCGAACAGGATATGCTGCTG	720
Q	721	CAACCGGGTAAAGCGCGCTGCATATGCAACCGCAAGCGGAGAGTATGACGTTACC	780
D	721	CAACCGGGTAAAGCGCGCTGCATATGCAACCGCAAGCGGAGAGTATGACGTTACC	780
Q	781	GATGCGGCGCAACGATGATTTGCGCTCTGCGCGCAACGCTGACAGCGGTAATTCGCTG	840
D	781	GATGCGGCGCAACGATGATTTGCGCTCTGCGCGCAACGCTGACAGCGGTAATTCGCTG	840
Q	841	GAAAGAGCTCTCTTCTTTCGCAATGCGCGCGCTGCGTGTGCTCGCAAACTGGGAAAC	900
D	841	GAAAGAGCTCTCTTCTTTCGCAATGCGCGCGCTGCGTGTGCTCGCAAACTGGGTAACA	900
Q	901	TTCACGGTTTCCGCAATCGAGTGAAGAAATGCTGTACGTGACGTCAGATACAGGCTT	960
D	901	TTCACGGTTTCCGCAATCGAGTGAAGAAATGCTGTACGTGACGTCAGATACAGGCTT	960
Q	961	GCGGTATGACCGAAGAGAACTGAAAGCTGCGGTGACGCGCAGCGCTAAACGTGTGAA	1020
D	961	GCGGTATGACCGAAGAGAACTGAAAGCTGCGGTGACGCGCAGCGCTAAACGTGTGAA	1020
Q	1021	AAAGTGTATGACCAACGATGCTTTGACATCTGACGCGCGGACGCTCTTATCTG	1080
D	1021	AAAGTGTATGACCAACGATGCTTTGACATCTGACGCGCGGACGCTCTTATCTG	1080
Q	1081	GCAAAATCCGCGAAGCGGTGACCGCTTATGTTGCGCTGCAACAGCGATGCTCCACC	1140
D	1081	GCAAAATCCGCGAAGCGGTGACCGCTTATGTTGCGCTGCAACAGCGATGCTCCACC	1140
Q	1141	AAACGCTGAAAGCGGAAACCGCGCGGTGAAACCGCTGGAACGCGGATGATGCTGCTG	1200
D	1141	AAACGCTGAAAGCGGAAACCGCGCGGTGAAACCGCTGGAACGCGGATGATGCTGCTG	1200
Q	1201	GCGGCACTGGAAGCGGTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1260
D	1201	GCGGCACTGGAAGCGGTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1260
Q	1261	ATGCGCGGATCTTTCGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1320
D	1261	ATGCGCGGATCTTTCGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1320
Q	1321	GCGGAGTAAAGAGTCTGCGCAACGCTGCGGAAAGTGTGTGCTCAACTTTGAAAGAC	1380
D	1321	GCGGAGTAAAGAGTCTGCGCAACGCTGCGGAAAGTGTGTGCTCAACTTTGAAAGAT	1380
Q	1381	GCTTGTGACGACCAATCATCAAGAGATCAACAGATA 1423	



Qy 982 CTGAAGTGGCCGTAGCGGACGCCGTAAACGTGTGTGAAAAAGTGTGATGACCAACGGT 1041  
Cc |||||  
Cc 955 TTGAGACAGGCGCGTCCGACGCGCGGTAAAGGTGGCGAGAAAGTGTGATACCAACGGC 1014  
Cc |||||  
Qy 1042 GTCTTTGACATCTGCAACGCGGCGACGTCTTTATCTGGCAAAATGCCCGACCTGGGT 1101  
Cc |||||  
Db 1015 GTTTTCGATATCTGCAACGCGGCGCACGTCTTATCTGGCGAAACGCGCGCAAACTGGGC 1074  
Cc |||||  
Qy 1102 GACCGCTGATTTGTGGCGGTCAACAGGATGCTCCACCAACGCGCTGAAAGGGGATTC 1161  
Cc |||||  
Db 1075 GACCGCTGATTTGTGGCGGTCAATATGACCGCTTCACTTAACGCTTGAAGGCGAAAGC 1134  
Cc |||||  
Qy 1162 CGCCCGGTAAACCACTGCAACAGCGTATGATGTGTGGCGCACTGAGACGGTGCAC 1221  
Cc |||||  
Db 1135 CGTCCGGTATTCGGCTGAAACAGGTATGATCGTGGCGCGGCTGGAGTCCGTGCAC 1194  
Cc |||||  
Qy 1222 TGGGTATGTGTGTTGAAGAGACACGCGCGGCTTATCGCGCGGATTTGCCAGAT 1281  
Cc |||||  
Db 1195 TGGATTTGTCTTTTGAAGAGATACCGCGCAACGACTGATTCGCGTATCTGCCGAT 1254  
Cc |||||  
Qy 1282 CTGCTGGTGAAGGCGCGCACTATTAACCAAGAGATTCGCGGAGTAAAGATCTGG 1341  
Cc |||||  
Db 1255 CTGCTGGTGAAGGCGCGCACTATTAACCGGAGATCGCGGACGCGAGAGGTCT-G 1313  
Cc |||||  
Qy 1342 GCCAACGGTGGCGAAGTGTGTGCTCAACTTTGAAGACGGTGTCTGACGACCAACATC 1401  
Cc |||||  
Db 1314 GCCAACGGCGCGCAAGTCACTGTGTGAACTTCGAAGATGTTGTTCCAGACCAATATTC 1373  
Cc |||||  
Qy 1402 ATCAAGAAATCCCAACAGATTA 1423  
Cc |||||  
Db 1374 ATCAAAAGATCCAGACCGAGA 1395  
Cc |||||

## RESULT 12

ID ACAS4111 standard; DNA; 1431 BP.  
XX ACAS4111,  
AC ACAS4111,  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE prokaryotic essential gene #35768.  
XX  
KW Antisense; db: prokaryotic essential gene; cell proliferation;  
XX drug design; gene.  
XX  
OS Versinia pectis.  
XX  
PN MO20027183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PE 21-MAR-2002; 2002MO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
XX 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Travick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
DR WPI; 2003-029926/02.  
DR P-PSDB; ABUS0241.  
XX  
XX New antisense nucleic acids, useful for identifying proteins or screening  
XX for homologous nucleic acids required for cellular proliferation to  
XX isolate candidate molecules for rational drug discovery programs.  
PS Claim 14; SEQ ID NO 41981; 1766bp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
Cc the 6213 antisense sequences given in the specification where expression  
Cc of the nucleic acid inhibits proliferation of a cell. Also included are:  
Cc (1) a vector comprising a promoter operably linked to the nucleic acid  
Cc encoding a polypeptide whose expression is inhibited by the antisense  
Cc nucleic acid; (2) a host cell containing the vector; (3) an isolated  
Cc polypeptide or its fragment whose expression is inhibited by the  
Cc antisense nucleic acid; (4) an antibody capable of specifically binding  
Cc the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
Cc proliferation or the activity of a gene in an operon required for  
Cc proliferation; (7) identifying a compound that influences the activity of  
Cc the gene product or that has an activity against a biological pathway  
Cc required for proliferation, or that inhibits cellular proliferation; (8)  
Cc identifying a gene required for cellular proliferation or the biological  
Cc pathway in which a proliferation-regulated gene or its gene product lies  
Cc or a gene on which the test compound that inhibits proliferation of an  
Cc organism acts; (9) manufacturing an antibiotic; (10) profiling a  
Cc compound's activity; (11) a culture comprising strains in which the gene  
Cc product is overexpressed or underexpressed; (12) determining the extent  
Cc to which each of the strains is present in a culture or collection of  
Cc strains; or (13) identifying the target of a compound that inhibits the  
Cc proliferation of an organism. The antisense nucleic acids are useful for  
Cc identifying proteins or screening for homologous nucleic acids required  
Cc for cellular proliferation to isolate candidate molecules for rational  
Cc drug discovery programs, or for screening homologous nucleic acids  
Cc required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
Cc *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
Cc prokaryotic essential genes. Note: The sequence data for this patent did  
Cc not form part of the printed specification, but was obtained in  
Cc electronic format directly from WIPO at  
Cc ftp.wipo.int/pub/published\_pct\_sequences  
XX

SO Sequence 1431 BP; 314 A; 325 C; 433 G; 359 T; 0 U; 0 Other;

Query Match 61.3%; Score 879.2; DB 8; Length 1431;  
Best Local Similarity 76.2%; Pred. No. 1,9e-250;  
Matches 1082; Conservative 0; Mismatches 338; Indels 0; Gaps 0;

Qy 1 ATGAAGTAAAGCTGCGACAGGTTTGAACGTCGAGAGTATGTTGGTGTGATGATG 60  
Cc |||||  
Db 1 ATGAAGTAAAGCTGCGACGCTGATTTTCCGCGTGTGATTAATCGTGGTATGATG 60  
Cc |||||  
Qy 61 CTGATCGTTACTGTGTAACGCGCCACACAGTGTATCTCGCCGGAAGCGCCGCGTG 120  
Cc |||||  
Db 61 TTAGACCGTACTGTATACGCGGCAACCTGCCGATTTACAGAGACCTGTGCTGTA 120  
Cc |||||  
Qy 121 GTTAAAGTAAATCCATCGAAGACGTCGCGCGCGCTTAAAGTGGGATGATATC 180  
Cc |||||  
Db 121 GTGAAGTCAATCAATCGAAGACGCGCTGTGTGTGCAAGCAACGTGGATGAATAT 180  
Cc |||||  
Qy 181 GCTTCTCGGATCTATATGACGCGTGTGCGGTTGACGGGCAATGACATGACGCGC 240  
Cc |||||  
Db 181 GCTTCTCGGATCTATATGACGCGTGTGCGGTTGACGGGCAATGACATGACGCGC 240  
Cc |||||  
Qy 241 GCGCTAGTAAATCTGTGCGGACGTCAACGTCAATGCGACTTCTTGTATCCGACG 300  
Cc |||||  
Db 241 GCGCTAGTCAAGGTCGAGCAAGTTCGGGTGCGTGGCATTTTGTCTCGTACCACT 300  
Cc |||||  
Qy 301 CATCCGACATTACCAATTAACGGTATCTTCCGCAACCAACGCTGATCCGTTCGAT 360  
Cc |||||  
Db 301 CACCCGACATTACCAATTAACGGTATCTTCCGCAACCAACGCTGATCCGTTCGAT 360  
Cc |||||  
Qy 361 TTTGAAGAAAGTTTCAAGGTTGATTCGACGACGCGTGCAGAGGCGATTAATACGG 420  
Cc |||||  
Db 361 TTTGAAGAAAGTTTCAAGGTTGATTCGACGACGCGTGCAGAGGCGATTAATACGG 420  
Cc |||||  
Qy 421 CTGATTCGATTTGCGCGCGTGTGCTTCTGACTACGCAAGGTGCGTGCAGAGGCTA 480  
Cc |||||  
Db 421 TTGCGACAGATTTGCGCGCGTGTGCTGCTGCAATTAAGCAAGGGGAGATTGAACAGG 480  
Cc |||||  
Qy 481 CAGCAATGATCAACTGCGCGGTAAAGCGGTTCCGGTCTGATGATCCAAAGGT 540  
Cc |||||

```

Db 481 CAGCCGATGATCCAACTGCGACGTAAGCCAAATGTCGGGTTTGTGATTCGAAAGGC 540
Qy 541 ACCGATTTTGAAGCGCTACCGCGCGCTACGCTGTTAAAGCGGAATCTCTCGAAATTTGAA 600
Db 541 AGTGAATTTGAGCGCTACCGGTGTCACATTTGTTGACGCGGAATTTGTCTGAATTTGAA 600
Qy 601 GCTGTTTCCGTAATTTGAAGCCGAAGAGATTTGTGAGCGCGCATGAAATCTGAT 660
Db 601 GCAATGATGAGCGCGCTGTAATAAATGAAGAGATTTGTTAAACGCGGATATGCAACTGAG 660
Qy 661 GCCGATTAACGAATCTCGGCTCTGTGTAGTGAACCGTTCCGAACAGGATATGCTGCTG 720
Db 661 GCAGATTTTGAATCTTGAAGATTTGCTGTGACCGGTTCTGAACAGGATATGACCTTTG 720
Qy 721 CAACCGGATTAAGCGCGCTGCAATATGCAACCCAGCGAGAGATGATGACGTTACC 780
Db 721 CAACGAGGTAACCGCGCTGCAATATGCAACCCAGCGAGAGATGATGACGTTACC 780
Qy 781 GGTGCGGCGGCAACGAGTGTGCTGTGCTGCGGCAACGCTGCGAGCGGGTAAATTCGCTG 840
Db 781 GGTGCTGCGGCAATGAGTGTGCTGTGCTGCGGCAACGCTGCGAGCGGGTAAATTCGCTG 840
Qy 841 GAAGAAGCTGCTCTTCTTTCGAATGCGGCGGCTGCGGCTGCGGCAACGCTGCGAACC 900
Db 841 GAAGAAGCTGCTCTTCTTTCGAATGCGGCGGCTGCGGCTGCGGCAACGCTGCGAACC 900
Qy 901 TCCAGGTTTTCGCGGATGAGCTGGAATGCTGTACGTGAGCGTGAATACAGGCTTT 960
Db 901 TCCAGGTTTTCGCGGATGAGCTGGAATGCTGTACGTGAGCGTGAATACAGGCTTT 960
Qy 961 GCGCTATGACCGGAAGGAATGAACTGACGCTGCGGCAACGCGGCTGAACGCTGGA 1020
Db 961 GCGCTATGACCGGAAGGAATGAACTGACGCTGCGGCAACGCGGCTGAACGCTGGA 1020
Qy 1021 AAGGTGATGATGACCAACGCTGCTTGAACATCTTGAACGCGCGGCAACGCTTATCTG 1080
Db 1021 AAGGTGATGATGACCAACGCTGCTTGAACATCTTGAACGCGCGGCAACGCTTATCTG 1080
Qy 1081 GCAAAATGCGCGGATGAGCTGCGGCTGCGGCTGCGGCTGCGGCAACGCTTATCTG 1140
Db 1081 GCAAAATGCGCGGATGAGCTGCGGCTGCGGCTGCGGCTGCGGCAACGCTTATCTG 1140
Qy 1141 AAGGTGATGATGACCAACGCTGCGGCTGCGGCTGCGGCTGCGGCAACGCTTATCTG 1200
Db 1141 AAGGTGATGATGACCAACGCTGCGGCTGCGGCTGCGGCTGCGGCAACGCTTATCTG 1200
Qy 1201 GCGGCACTGGAAGCGGCTGAGCTGAGTGTGCTTGAAGAGACAGCGCGGCTG 1260
Db 1201 GCGGCACTGGAAGCGGCTGAGCTGAGTGTGCTTGAAGAGACAGCGCGGCTG 1260
Qy 1261 ATGCGCGGATCTTGAAGATCTGCTGCTGAGTGTGAGAGCGCGGCTGATTAACAGAGAT 1320
Db 1261 ATGCGCGGATCTTGAAGATCTGCTGCTGAGTGTGAGAGCGCGGCTGATTAACAGAGAT 1320
Qy 1321 GCGGCACTGGAAGCGGCTGAGCTGAGTGTGCTTGAAGAGACAGCGCGGCTG 1380
Db 1321 GCGGCACTGGAAGCGGCTGAGCTGAGTGTGCTTGAAGAGACAGCGCGGCTG 1380
Qy 1381 GGTGCTGAGACCAACATCATCAAGAGATTCACAGG 1420
Db 1381 GGTGCTGAGACCAACATCATCAAGAGATTCACAGG 1420

```

RESULT 13  
ACH96312/c  
ID ACH96312 standard; DNA; 1323 BP.

ACH96312;  
29-JUL-2004 (first entry)  
Klebsiella pneumoniae polynucleotide seqid 2107.

```

KW Recombinant expression vector; transcription regulatory element;
KW Klebsiella pneumoniae protein; antibacterial; Vaccine; gene; ds.
XX Klebsiella pneumoniae.
XX US6610836-B1.
XX 26-AUG-2003.
XX 27-JAN-2000; 2000US-00489039.
XX 29-JAN-1999; 99US-0117747P.
XX (GENO-) GENOME THERAPEUTICS CORP.
XX Breton GL, Osborne M;
XX WPI; 2003-895346/82.
XX P-PSDB; ABO62761.
XX New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for
XX preparing a vaccine composition against Klebsiella pneumoniae.
XX Disclosure; SEQ ID NO 2107; 932pp; English.
XX The invention describes a new isolated nucleic acid encoding a Klebsiella
XX pneumoniae polypeptide. Also described are: a recombinant expression
XX vector comprising the nucleic acid, operably linked to a transcription
XX regulatory element; and a cell comprising the recombinant expression
XX vector. The nucleic acid is useful for preparing a vaccine composition
XX against Klebsiella pneumoniae. This sequence encodes a Klebsiella
XX pneumoniae polypeptide of the invention
XX
SQ Sequence 1323 BP; 244 A; 450 C; 376 G; 253 T; 0 U; 0 Other;

```

```

Query Match 60.7%; Score 870.6; DB 11; Length 1323;
Best Local Similarity 81.8%; Pred. No. 6.6e-248;
Matches 1005; Conservative 0; Mismatches 224; Indels 0; Gaps 0;

```

```

Qy 1 ATGAAAGTAAACGCTGCGCAAGTGTGAACGTCAGAGATGATGCTGATGATG 60
Db 1230 ATGAAAGTAAACGCTGCGCAAGTGTGAACGTCAGAGATGATGCTGATGATG 1171
Qy 61 CTGATGCTTACTGATGAGCGGCGGCAACGCTGATGCTGCTGCGGCAACGCTGCGG 120
Db 1170 CTGATGCTTACTGATGAGCGGCGGCAACGCTGATGCTGCTGCGGCAACGCTGCGG 1111
Qy 121 GTTAAAGTAAACGCTGCGCAAGTGTGAACGTCAGAGATGATGCTGATGATG 180
Db 1110 GTTAAAGTAAACGCTGCGCAAGTGTGAACGTCAGAGATGATGCTGATGATG 1051
Qy 181 GCTTCTCTCGGCTGATGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
Db 1050 GCTTCTCTCGGCTGATGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 991
Qy 241 GCGCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
Db 990 GCGCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 931
Qy 301 CATCGACATTAACCAATTAAGGATGATGCTTCCGCAACCAAGCTGATCCGCTGAT 360
Db 930 CATCGACATTAACCAATTAAGGATGATGCTTCCGCAACCAAGCTGATCCGCTGAT 871
Qy 361 TTGAAGAGATTTGAAGAGTGTGATTCGCAACGCGCTGCAAGCGGATTAACAGGCG 420
Db 870 TTGAAGAGATTTGAAGAGTGTGATTCGCAACGCGCTGCAAGCGGATTAACAGGCG 811
Qy 421 CTGATGCTTACTGATGAGCGGCGGCAACGCTGATGCTGCTGCTGCTGCTGCTGCTG 480
Db 810 CTGATGCTTACTGATGAGCGGCGGCAACGCTGATGCTGCTGCTGCTGCTGCTGCTG 751
Qy 481 CAGCAATGATCAACGCTGCGGCTGAAGCGGATGCTGCTGCTGCTGCTGCTGCTGCTG 540

```

```

Db 750 CAGACATGATCCGACTCGACGTAAGCGGCGTGCCTGATCATCCGAAAGGC 691
Qy 541 ACCGATTTTGAGCGCTACCGGCGGCTACGCTGTTAAACCGGAATCTCTGGAAATTTGAA 600
Db 690 ACCGATTTTGAGCGCTACCGGCGGCTACGCTGTTAAACCGGAATCTCTGGAAATTTGAA 631
Qy 601 GCTGTGTGCTGTAATGATGACCGAAGAAAGATTTGTAAGCGCGGCAATGAATCTGATTT 660
Db 630 GCGGTGGTGGCAAGTGCAGAGATGAAGCGAGATCGTTGAGCCGGGATGAATCATTC 571
Qy 661 GCCGATTAACGAATCTCGGCTCTGTAGTGAACCGGTTCCGAACAGGGATATGCTGCTG 720
Db 570 GCCGAATTCGAATCTCGGCTCTGTAGTGAACCGGTTCCGAACAGGGATATGCTGCTG 511
Qy 721 CAACCGGGTAAAGCGCGCTGATATGCAACCAAGCGGAGAGATGATAGCTTACC 780
Db 510 CAGCGGGAGCTCGGCGCTGATATGCAACCAAGCGGAGAGATGATAGCTTACC 451
Qy 781 GGTGCGGCGCAACAGGTGATTTGGCTCTGCGGCAACGCTGAGCGGATTTGCTG 840
Db 450 GGGCGCGGCGCAACAGGTGATTTGGCTCTGCGGCAACGCTGAGCGGATTTGCTG 391
Qy 841 GAAGAAGCTCTCTTTTTCGCAATGCGCGGCTGCGTGGTGTGCGCAACCTGGAAAC 900
Db 390 GAAGAAGCTCTCTTTTTCGCAATGCGCGGCTGCGTGGTGTGCGCAACCTGGAAAC 331
Qy 901 TCACGCGTTTCCGCGATCGAGCTGGAATAATGCTGATCGTGAACGATGACAGGCTTT 960
Db 330 TCACGCGTTTCCGCGATCGAGCTGGAATAATGCTGATCGTGAACGATGACAGGCTTT 271
Qy 961 GGCCTGATGACCGAAGAACTGAAAGTGGCGGTAGCGGAGCGGCTAAACGTGTGAA 1020
Db 270 GGCCTGATGACCGAAGAACTGAAAGTGGCGGTAGCGGAGCGGAGCGGAG 211
Qy 1021 AAATGTGTGATGACCAAGGTGTCTTTGACATCTCTGACGCGCGGCACTCTTATCTG 1080
Db 210 AAATGTGTGATGACCAAGGTGTCTTTGACATCTCTGACGCGGTCTCTTATCTG 151
Qy 1081 GCAATATGCGCGCAAGCTGAGTGAACGCTGATTTGTCGCTGCAACAGGATGCTTCAAC 1140
Db 150 GCGAATGCGCGCAAGCTGAGTGAACGCTGATTTGTCGCTGCAACAGGATGCTTCAAC 91
Qy 1141 AAACGCTGAAAGGAGATTTCCGCGCTGAAACCACTCGAAGCGGTATGATTTGCTG 1200
Db 90 AAACGCTGAAAGGAGATTTCCGCGCTGAAACCACTCGAAGCGGTATGATTTGCTG 31
Qy 1201 GCGCAGCTGAAAGGAGTGTGATGCTGCTG 1229
Db 30 GGGCGCTGAAAGGAGTGTGATGCTGCTG 2

```

## RESULT 14

ACF67367\_45  
Continuation (46 of 57) of ACF67367 from base 450001 (Photobacterium luminescens nucleotid)

Fragment Name Begin End  
WP ACF67367\_00 1 110000  
WP ACF67367\_01 100001 210000  
WP ACF67367\_02 200001 310000  
WP ACF67367\_03 300001 410000  
WP ACF67367\_04 400001 510000  
WP ACF67367\_05 500001 610000  
WP ACF67367\_06 600001 710000  
WP ACF67367\_07 700001 810000  
WP ACF67367\_08 800001 910000  
WP ACF67367\_09 900001 1010000  
WP ACF67367\_10 1000001 1110000  
WP ACF67367\_11 1100001 1210000  
WP ACF67367\_12 1200001 1310000  
WP ACF67367\_13 1300001 1410000  
WP ACF67367\_14 1400001 1510000  
WP ACF67367\_15 1500001 1610000  
WP ACF67367\_16 1600001 1710000

```

WP ACF67367_17 1700001 1810000
WP ACF67367_18 1800001 1910000
WP ACF67367_19 1900001 2010000
WP ACF67367_20 2000001 2110000
WP ACF67367_21 2100001 2210000
WP ACF67367_22 2200001 2310000
WP ACF67367_23 2300001 2410000
WP ACF67367_24 2400001 2510000
WP ACF67367_25 2500001 2610000
WP ACF67367_26 2600001 2710000
WP ACF67367_27 2700001 2810000
WP ACF67367_28 2800001 2910000
WP ACF67367_29 2900001 3010000
WP ACF67367_30 3000001 3110000
WP ACF67367_31 3100001 3210000
WP ACF67367_32 3200001 3310000
WP ACF67367_33 3300001 3410000
WP ACF67367_34 3400001 3510000
WP ACF67367_35 3500001 3610000
WP ACF67367_36 3600001 3710000
WP ACF67367_37 3700001 3810000
WP ACF67367_38 3800001 3910000
WP ACF67367_39 3900001 4010000
WP ACF67367_40 4000001 4110000
WP ACF67367_41 4100001 4210000
WP ACF67367_42 4200001 4310000
WP ACF67367_43 4300001 4410000
WP ACF67367_44 4400001 4510000
WP ACF67367_45 4500001 4610000
WP ACF67367_46 4600001 4710000
WP ACF67367_47 4700001 4810000
WP ACF67367_48 4800001 4910000
WP ACF67367_49 4900001 5010000
WP ACF67367_50 5000001 5110000
WP ACF67367_51 5100001 5210000
WP ACF67367_52 5200001 5310000
WP ACF67367_53 5300001 5410000
WP ACF67367_54 5400001 5510000
WP ACF67367_55 5500001 5610000
WP ACF67367_56 5600001 5648894

```

Query Match 54.7%; Score 784.8; DB 10; Length 110000;

Best Local Similarity 71.8%; Pred. No. 1,8e-221; Mismatches 402; Indels 0; Gaps 0;

Matches 1026; Conservative 0;

```

Qy 1 ATGAAGTAAAGCTGCGAGAGTTGAACGTCAGAGTATGCTGCTGAGCGGCTGATG 60
Db 38374 ATGAAGTAAAGCTGCGAGAGTTGAACGTCAGAGTATGCTGCTGAGCGGCTGATG 38433
Qy 61 CTGATGCTGATGTAAGCGGCGCCACAGCTGATCTCGCGGAGCGCGGCTGCTG 120
Db 38434 TTGATGCTGATGTAAGCGGCGCCACAGCTGATCTCGCGGAGCGCGGCTGCTG 38493
Qy 121 GTTAAAGTAAAGCTGCGAGAGTTGAACGTCAGAGTATGCTGCTGAGCGGCTGATG 180
Db 38494 GTTAAAGTAAAGCTGCGAGAGTTGAACGTCAGAGTATGCTGCTGAGCGGCTGATG 38553
Qy 181 GCTTCTGCTGCTGATGTAAGCGGCGCCACAGCTGATCTCGCGGAGCGCGGCTGCTG 240
Db 38554 GCTTCTGCTGCTGATGTAAGCGGCGCCACAGCTGATCTCGCGGAGCGCGGCTGCTG 38613
Qy 241 GCGCTGATGTAAGCTGCGAGAGTTGAACGTCAGAGTATGCTGCTGAGCGGCTGATG 300
Db 38614 GCGCTGATGTAAGCTGCGAGAGTTGAACGTCAGAGTATGCTGCTGAGCGGCTGATG 38673
Qy 301 CATTCGACCAATTCGTAAGCGGCTGATCTCGCGGAGCGCGGCTGATCTGATG 360
Db 38674 CATTCGACCAATTCGTAAGCGGCTGATCTCGCGGAGCGCGGCTGATCTGATG 38733
Qy 361 TTGAAGAGGTTTGAAGGTTGATGCGAGCGGCTGATCTCGCGGAGCGCGGCTGATG 420
Db 38734 TTGAAGAGGTTTGAAGGTTGATGCGAGCGGCTGATCTCGCGGAGCGCGGCTGATG 38793

```

Qy 421 CTGAGTTGATTCGCGGCTGCTGCTTCTGACTACGCAAAAGGCTGGCAAGCGTA 480  
 Db 38794 TTACCCATATTTGAGACATTTGTTCTATCTGACTACGCAAAAGGCTGCACTGAGTCAGGTT 38853  
 Qy 481 CAGCAGATGATTCGACTGCGGCGGTAAAGCGGGTCTCCGCTGCTGATTTGATTCGCAAAAGT 540  
 Db 38854 CAGGAAATGATTCGAACTGGCAAAATCGCGGCAAAAGTCCCGGATTTAATTCGATTCGAAAGT 38913  
 Qy 541 ACCGATTTTGAAGCGCTACCGCGCGCTACCGCTGTTAAACCGCAATCTCTCGAAATTTGAA 600  
 Db 38914 AATGATTTTGAAGCGCTTACGAGGTGCAACTTGTCTAAACCGCAATCTCTCGAAATTTGAA 38973  
 Qy 601 GCTGTGCTGCTTAAATGTAAGACCGAAGAAAGATTTGAGCGCGGATGAAATCTGATT 660  
 Db 38974 GCGGTCTGCGGCTACCTGCAAAAGATGCAATGAGTTGTTGAAAGGAACTCGGCTAGTT 39033  
 Qy 661 GCCGATTAAGAACTCTGCGCTGCTGTTAGTACCGGCTTCGAAACGAGGTATGTCGCTG 720  
 Db 39034 AAGGATTTGATCTTCAAGGCACTTACTGATCAACCGCTTCGAAACGAGGATGAGTTGCTG 39093  
 Qy 721 CAACCGGTTAAAGCGCGCTGCTGATATGCAACCGCAAGAGTGTATGAGTTACC 780  
 Db 39094 AGTGTGATCAAGCCACATGCACTTGTCTACGAGCGCGCAAGAGTGTATGAGTTACG 39153  
 Qy 781 GGTGCGGCGCAACAGGTGATTTGCGCTCTGCGCGCAACGCTGCGAGCGGTAATTGCTG 840  
 Db 39154 GGTGCTGCGCAATACAGTTATTTGCTGTGTTACCAACGCGGATTTGCTGCGAAACCACTG 39213  
 Qy 841 GAAGAACCTGCTTCTTTCGCAATGCGCGCGCTGCGCTGCTGCTGCGCAACGCGGAAC 900  
 Db 39214 AATGAACTTGTCTTCTGCGCAATGCTGCGCGGCTGCTGCTGCTGCTGCAACGCGGAAC 39273  
 Qy 901 TCCACGCTTTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960  
 Db 39274 TCGACCTTTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 39333  
 Qy 961 GCGCTGATGACCGAAGGAACTGAGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020  
 Db 39334 GGTGTGATGACCGAATTTCACTTAAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 39393  
 Qy 1021 AAGGTGATGACCAACGCTGCTTGTGCACTCTGCAACGCGCGCACTCTCTTATCTG 1080  
 Db 39394 AGGCTGATGACTAATGCTGCTTGTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 39453  
 Qy 1081 GCAAATCCCGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140  
 Db 39454 GAGAAATCCCGTGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 39513  
 Qy 1141 AAGCGCTGAAGGCGATTCGCGCGGTAACCACTCGAAGCGGTAATGATTTGCTG 1200  
 Db 39514 AAGCGCTGAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 39573  
 Qy 1201 GCGGCACTGAAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260  
 Db 39574 TCGCTCTGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 39633  
 Qy 1261 ATGCGCGGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320  
 Db 39634 ATGCGCGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 39693  
 Qy 1321 GCGCGGATTAAGAAAGTCTGCGCGCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380  
 Db 39694 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 39753  
 Qy 1381 GGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1428  
 Db 39754 GGTATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 39811

RESULT 15  
 ACP65388\_02/c  
 Continuation (3 of 13) of ACP65388 from base 200001 (Photobacterium luminescens nucleotide  
 WP Sequence split into 13 fragments LOCUS ACP65388 Accesion ACP65388

WP Fragment Name Begin End  
 WP ACP65388\_01 1 110000  
 WP ACP65388\_02 200001 210000  
 WP ACP65388\_03 200001 310000  
 WP ACP65388\_04 400001 410000  
 WP ACP65388\_05 400001 510000  
 WP ACP65388\_06 500001 610000  
 WP ACP65388\_07 600001 710000  
 WP ACP65388\_08 700001 810000  
 WP ACP65388\_09 800001 910000  
 WP ACP65388\_10 900001 1010000  
 WP ACP65388\_11 1000001 1110000  
 WP ACP65388\_12 1200001 1225559

Query Match 54.7%; Score 784.8; DB 10; Length 110000;  
 Best Local Similarity 71.8%; Pred. No. 1,8e-221;  
 Matches 1026; Conservative 0; Mismatches 402; Indels 0; Gaps 0;

Qy 1 ATGAAAGTAAAGCTGCGAGGTTTGAACGTCAGAGTGAATGCTGCTGCTGCTGCTGCTG 60  
 Db 86563 ATGAAAGTAAAGCTGCGAGGTTTGAACGTCAGAGTGAATGCTGCTGCTGCTGCTGCTG 86504  
 Qy 61 CTGATCGTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120  
 Db 86503 TTGATCGTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 86444  
 Qy 121 GTTAAAGTAAAGCTGCGAGGTTTGAACGTCAGAGTGAATGCTGCTGCTGCTGCTGCTG 180  
 Db 86443 GTTAAAGTAAAGCTGCGAGGTTTGAACGTCAGAGTGAATGCTGCTGCTGCTGCTGCTG 86384  
 Qy 181 GCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240  
 Db 86383 GCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 86324  
 Qy 241 GCGCTGATTAATCTCTGCGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300  
 Db 86323 GCGCTGATTAATCTCTGCGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 86264  
 Qy 301 CATCCGACATTAACCAATTAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360  
 Db 86263 CATCCGACATTAACCAATTAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 86204  
 Qy 361 TTTGAAAGGTTTGAAGGTTTGAAGGTTTGAAGGTTTGAAGGTTTGAAGGTTTGAAGGTT 420  
 Db 86203 TTTGAAAGGTTTGAAGGTTTGAAGGTTTGAAGGTTTGAAGGTTTGAAGGTTTGAAGGTT 86144  
 Qy 421 CTGAGTTGATTCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480  
 Db 86143 TTACCCATATTTGAGACATTTGTTCTATCTGACTACGCAAAAGGCTGCACTGAGTCAGGTT 86084  
 Qy 481 CAGCAGATGATTCGACTGCGGCGGTAAAGCGGGTCTCCGCTGCTGATTTGATTCGAAAGT 540  
 Db 86083 CAGCAGATGATTCGACTGCGGCGGTAAAGCGGGTCTCCGCTGCTGATTTGATTCGAAAGT 86024  
 Qy 541 ACCGATTTTGAAGCGCTACCGCGCGCTACCGCTTAAACCGCAATCTCTCGAAATTTGAA 600  
 Db 86023 AATGATTTTGAAGCGCTTACGAGGTGCACTTGTCTAAACCGCAATCTCTCGAAATTTGAA 85964  
 Qy 601 GCTGTGCTGCTTAAATGTAAGACCGAAGAAAGATTTGAGCGCGGATGAAATCTGATT 660  
 Db 85963 GCGGTCTGCGGCTACCTGCAAAAGATGCAATGAGTTGTTGAAAGGAACTCGGCTAGTT 85904  
 Qy 661 GCCGATTAAGAACTCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720  
 Db 85903 AAGGATTTGATCTTCAAGGATTAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 85844  
 Qy 721 CAACCGGTTAAAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780  
 Db 85843 AGTGTGATGACGCACTGCACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 85784  
 Qy 781 GGTGCGGCGCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840



Db 85783 GGTGCTGGCGATACGATTATGTGTGTAGCAACGGCGATTGCGGCTGGAAGCCACTG 85724  
QY 841 GAAGAAGCCCTGCTCTTTGCCAATGCGCGGCTGGCGTGGTGGCGAACTGGGAACC 900  
Db 85723 AATGAAGCTTGTTCCTGCGCAATGCTGCGGGGTTGTTGGCAACTGGGAACG 85664  
QY 901 TCCACGGTTTCGCCGATCGAGCTGAAAATGCTGATGAGCGTGCAAGTACAGGCTTT 960  
Db 85663 TCGACGGTTTCGCCGATTGAACTGGAAGATGCTATCCGTGGCCGCGTGAACAGGTTTC 85604  
QY 961 GGCCTGATGACCGAAGAGAACTGAAGTGGCCGTAGCGGCAAGCGCTMAAAGTGTGAA 1020  
Db 85603 GGTGTGATGACCGAATTCAAGCTTAAACAGGCTGTGTGATGCTCTCAACGTGTGAA 85544  
QY 1021 AAAGTGTGATGACCAACGGTGTCTTGAACATCTGCAACGCCGGCAAGTCTTATCTG 1080  
Db 85543 AGGCTGTGATGACTTAATGCTGCTTGAATTTCTGCAAGCAGGTCAATGCTCTTATCTG 85484  
QY 1081 GCAATGCGCGCAAGCTGGTGAACGCTTGAATGTTGCCGTCAACAGCGATGCTCCACC 1140  
Db 85483 GAGAAATGCCGTAGACTGGGTGATCGGTTAATGTTGCTGTTACAGTATGATCAACT 85424  
QY 1141 AAACGCTGAAAAGGGGATTCGCCGCCGTTAAACCACTCGAACAGCTATGATGTGCTG 1200  
Db 85423 AAACGCTGAAAAGGTGAGAGTCTGCCAGTTAATCAGTAAACAGAAATGATGTGTTA 85364  
QY 1201 GCGGCACTGGAAGGGGTGCACTGGGTAGTGTGTTTGAAGAGSACAGCGCGCAGGCTTG 1260  
Db 85363 TCTGCTGCGGCGCGGTGATGAGTGTGTTCCATTGAGAGATPACACACAGCGAATG 85304  
QY 1261 ATCCCGGGAATCTTGCCAGATCTGTGTTGAAGCGCGCACTATMAACGAGAGATT 1320  
Db 85303 ATTGCGGATGTTTTGCCGGAATGTAATGGTAAAGGGGAGATTATMAACCGGAGAAATT 85244  
QY 1321 GCCGGAGTAAAGAAAGTCTGGGCCAAGCTGGCGAAGTGTGCTCACTTTGAAGAC 1380  
Db 85243 GCTGTAGTGAAGAGTTTGGGCTGCGGGCGAGACGTCAAAGTGTGAACCTTGAAGAT 85184  
QY 1381 GGTGCTGACGACCAATCATCATAGAAAGATCCACAGATTAATAA 1428  
Db 85183 GGTATTTGACACCAATATCATCAAGGCGATTMAAATCAGTAAATA 85136

Search completed: March 18, 2006, 19:07:31  
Job time : 945 secs





inhibition occurs in the second microorganism. The nucleic acid sequences identified as being required for bacterial growth and proliferation, can be used for antisense therapy for killing bacteria

Sequence 477 AA:

Query Match 100.0%; Score 2393; DB 3; Length 477;  
Best Local Similarity 100.0%; Pred. No. 6.9e-207;  
Matches 477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKVTLPFERBAGVWVGVMDLDRYWGPTSRISPEAPVPVVKVNTIEERPGGANVAMNI 60  
DB 1 MKVTLPFERBAGVWVGVMDLDRYWGPTSRISPEAPVPVVKVNTIEERPGGANVAMNI 60  
QY 61 ASLGANARLVGLTGIDDAARALSKSLADVANKCPVSPVPTTKLRVLSRNOQLRLD 120  
DB 61 ASLGANARLVGLTGIDDAARALSKSLADVANKCPVSPVPTTKLRVLSRNOQLRLD 120  
QY 121 FEEGFEVDPOPLHERINQALSSIGALVLSDYAKGALASVOQMIQLARKAGVPVLIDPKG 180  
DB 121 FEEGFEVDPOPLHERINQALSSIGALVLSDYAKGALASVOQMIQLARKAGVPVLIDPKG 180  
QY 181 TDFERYRGATLTTPNLSFEFAVVGKCTEERIVERGKMLADYELSLALVTRSEQMSLL 240  
DB 181 TDFERYRGATLTTPNLSFEFAVVGKCTEERIVERGKMLADYELSLALVTRSEQMSLL 240  
QY 241 QPGKAPLHPTQAOEYVDVAGDVTYIGVLAATLAAGNSLEBACFPANAAAGVVGKLG 300  
DB 241 QPGKAPLHPTQAOEYVDVAGDVTYIGVLAATLAAGNSLEBACFPANAAAGVVGKLG 300  
QY 301 STVSPTELENAVGRADTGFVMTBEELKLAVAAARKGKRVMTNGVFDILAHGVSYL 360  
DB 301 STVSPTELENAVGRADTGFVMTBEELKLAVAAARKGKRVMTNGVFDILAHGVSYL 360  
QY 361 ANARKLGDRLIVAVNSDASTKRLKGSRPVNPLEORMIVLGALRAVDMVVSFEEDTQRL 420  
DB 361 ANARKLGDRLIVAVNSDASTKRLKGSRPVNPLEORMIVLGALRAVDMVVSFEEDTQRL 420  
QY 421 IAGILPDLVKGDDYKPEEIASGKEVMANGSEVLVNFEDGCSITNIIKKIQODKKG 477  
DB 421 IAGILPDLVKGDDYKPEEIASGKEVMANGSEVLVNFEDGCSITNIIKKIQODKKG 477

RESULT 2  
ID AU34708 standard; protein; 477 AA.

AC AU34708;

14-FEB-2002 (first entry)

E. coli cellular proliferation protein #289.

Antisense; prokaryotic cellular proliferation protein; antibiotic;  
antibacterial; drug design.

Escherichia coli.

MO200170955-A2.

27-SEP-2001.

21-MAR-2001; 2001WO-US009180.

21-MAR-2000; 2000US-0191078P.

23-MAY-2000; 2000US-0206848P.

26-MAY-2000; 2000US-0207727P.

23-OCT-2000; 2000US-0242578P.

27-NOV-2000; 2000US-0253625P.

22-DEC-2000; 2000US-0257931P.

16-FEB-2001; 2001US-0269308P.

XX Haselbeck R, Ohlsen KL, Zykkind JW, Wall D, Trawick JD, Carr GJ;  
PI Yamamoto RT, Xu HH;  
XX WPI, 2001-611495/70.  
DR N-PSDB; AAS52567.  
XX New polynucleotides for the identification and development of  
PT antibiotics, comprise sequences of antisense nucleic acids.  
XX

PS Example 3; SEQ ID NO 10301; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the genes,  
CC their use in the discovery of novel antibiotics, the essential genes  
CC themselves and the encoded proteins. The prokaryotes used are Escherichia  
CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,  
CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also  
CC useful for the identification of potential new targets for antibiotic  
CC development. The antisense nucleic acids can also be used to identify  
CC proteins used in proliferation, to express these proteins, and to obtain  
CC antibodies capable of binding to the expressed proteins. The proteins can  
CC be used to screen compounds in rational drug discovery programmes. The  
CC antisense nucleic acid sequence is also useful to screen for homologous  
CC nucleic acids which are required for cell proliferation in a wide variety  
CC of organisms. The present sequence represents an essential prokaryotic  
CC cellular proliferation protein. Note: The sequence data for this patent  
CC did not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 477 AA:

Query Match 100.0%; Score 2393; DB 4; Length 477;  
Best Local Similarity 100.0%; Pred. No. 6.9e-207;  
Matches 477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKVTLPFERBAGVWVGVMDLDRYWGPTSRISPEAPVPVVKVNTIEERPGGANVAMNI 60  
DB 1 MKVTLPFERBAGVWVGVMDLDRYWGPTSRISPEAPVPVVKVNTIEERPGGANVAMNI 60  
QY 61 ASLGANARLVGLTGIDDAARALSKSLADVANKCPVSPVPTTKLRVLSRNOQLRLD 120  
DB 61 ASLGANARLVGLTGIDDAARALSKSLADVANKCPVSPVPTTKLRVLSRNOQLRLD 120  
QY 121 FEEGFEVDPOPLHERINQALSSIGALVLSDYAKGALASVOQMIQLARKAGVPVLIDPKG 180  
DB 121 FEEGFEVDPOPLHERINQALSSIGALVLSDYAKGALASVOQMIQLARKAGVPVLIDPKG 180  
QY 181 TDFERYRGATLTTPNLSFEFAVVGKCTEERIVERGKMLADYELSLALVTRSEQMSLL 240  
DB 181 TDFERYRGATLTTPNLSFEFAVVGKCTEERIVERGKMLADYELSLALVTRSEQMSLL 240  
QY 241 QPGKAPLHPTQAOEYVDVAGDVTYIGVLAATLAAGNSLEBACFPANAAAGVVGKLG 300  
DB 241 QPGKAPLHPTQAOEYVDVAGDVTYIGVLAATLAAGNSLEBACFPANAAAGVVGKLG 300  
QY 301 STVSPTELENAVGRADTGFVMTBEELKLAVAAARKGKRVMTNGVFDILAHGVSYL 360  
DB 301 STVSPTELENAVGRADTGFVMTBEELKLAVAAARKGKRVMTNGVFDILAHGVSYL 360  
QY 361 ANARKLGDRLIVAVNSDASTKRLKGSRPVNPLEORMIVLGALRAVDMVVSFEEDTQRL 420  
DB 361 ANARKLGDRLIVAVNSDASTKRLKGSRPVNPLEORMIVLGALRAVDMVVSFEEDTQRL 420  
QY 421 IAGILPDLVKGDDYKPEEIASGKEVMANGSEVLVNFEDGCSITNIIKKIQODKKG 477  
DB 421 IAGILPDLVKGDDYKPEEIASGKEVMANGSEVLVNFEDGCSITNIIKKIQODKKG 477

RESULT 3  
ID AAG98421 standard; protein; 477 AA.

XX AC AAG98421;  
XX 21-SEP-2001 (first entry)  
XX Becherichia coli protein sequence SEQ ID NO:469.  
XX Becherichia coli; identification; proliferation; microorganism;  
XX antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;  
XX bacterial growth inhibition.  
XX Becherichia coli.  
XX MO200148209-A2.  
XX 05-JUL-2001.  
XX 19-DEC-2000; 2000MO-US034419.  
XX 23-DEC-1999; 99US-0173005P.  
XX (ELIT-) ELITRA PHARM INC.  
XX Foreych RA, Ohlsen KU, Zyskind JW;  
XX WPI, 2001-457376/49.  
XX N-PSDB; AAH81477.  
XX Novel nucleic acids encoding proteins required for Becherichia coli  
XX proliferation, useful for screening for antimicrobial agents.  
XX Example 4; Page 585-586; 596pp; English.  
XX The present invention describes a purified or isolated nucleic acid  
XX sequence (1) consisting essentially of one of the 93 nucleotide sequences  
XX given in AAH81202 to AAH81294, where expression of the nucleic acid in a  
XX microorganism is capable of inhibiting proliferation of a microorganism.  
XX (1) have antibacterial and antibiotic activities, and can be used in gene  
XX therapy. Expression of (1) in a microorganism inhibits proliferation of  
XX the microorganism, and the manufactured antibiotic is useful for reducing  
XX the activity or level of a gene product required for proliferation of a  
XX microorganism in a subject, specifically humans. The nucleic acids that  
XX inhibit bacterial growth or proliferation can be used as antisense  
XX therapeutics for killing bacteria. In addition to therapeutic  
XX applications, the nucleic acid sequences complementary to sequences  
XX required for proliferation can be used as diagnostic tools. For example,  
XX nucleic acid probes complementary to proliferation-required sequences  
XX that are specific for particular species of microorganisms can be used as  
XX probes to identify particular microorganism species in clinical  
XX specimens. AAH81295 to AAH81487 encode the Becherichia coli proteins  
XX given in AAG98239 to AAG98431, and AAH81488 to AAH81491 represent  
XX oligonucleotides, which are used in the exemplification of the present  
XX invention  
XX  
XX Sequence 477 AA;  
XX  
XX Query Match 100.0%; Score 2393; DB 4; Length 477;  
XX Best Local Similarity 100.0%; Pred. No. 6.9e-207;  
XX Matches 477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 181 TDPERYRGATLTPNLSFPAVVGKCTBEIYERGMKLIADVELSLVTRSEQMSLL 240  
XX  
XX 241 QPKAPLHMPDQAEVYDVTGADTVIGVLAATLAAGNSIEBACFPANAAAGVVGGLGT 300  
XX 241 QPKAPLHMPDQAEVYDVTGADTVIGVLAATLAAGNSIEBACFPANAAAGVVGGLGT 300  
XX 301 STSPSIFLENAVGRADTGGVWTEERELKLAVAARKRGKVMYMTGVPDILAHGVSYL 360  
XX 301 STSPSIFLENAVGRADTGGVWTEERELKLAVAARKRGKVMYMTGVPDILAHGVSYL 360  
XX 361 ANARKGDRILIVANSDSASTKRLKGSRPVNPPEORIVTGALBAVDWVVSFEEDIPORL 420  
XX 361 ANARKGDRILIVANSDSASTKRLKGSRPVNPPEORIVTGALBAVDWVVSFEEDIPORL 420  
XX 421 IAGILPDLVKGGDYKPEBIAGSKVWANGEVLVNFBEDGCTTNIKKIQDDKKG 477  
XX 421 IAGILPDLVKGGDYKPEBIAGSKVWANGEVLVNFBEDGCTTNIKKIQDDKKG 477  
XX  
XX RESULT 4  
XX ABU14690  
XX ID ABU14690 standard; protein; 477 AA.  
XX  
XX ABU14690;  
XX  
XX 19-JUN-2003 (first entry)  
XX  
XX Protein encoded by prokaryotic essential gene #217.  
XX  
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX  
XX Becherichia coli.  
XX  
XX MO200277183-A2.  
XX 03-OCT-2002.  
XX  
XX 21-MAR-2002; 2002MO-US009107.  
XX  
XX 21-MAR-2001; 2001US-00815242.  
XX 06-SEP-2001; 2001US-00948993.  
XX 25-OCT-2001; 2001US-0342923P.  
XX 08-FEB-2002; 2002US-00072851.  
XX 06-MAR-2002; 2002US-0362699P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
XX  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KU, Zyskind JW;  
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Foreych RA, Xu HH;  
XX WPI, 2003-029926/02.  
XX N-PSDB; ACA18560.  
XX  
XX New antisense nucleic acids, useful for identifying proteins or screening  
XX PT for homologous nucleic acids, required for cellular proliferation to  
XX PT isolate candidate molecules for rational drug discovery programs.  
XX  
XX Claim 25; SEQ ID NO 42614; 1766pp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising any one of  
XX the 6213 antisense sequences given in the specification where expression  
XX of the nucleic acid inhibits proliferation of a cell. Also included are:  
XX (1) a vector comprising a promoter operably linked to the nucleic acid  
XX encoding a polypeptide whose expression is inhibited by the antisense  
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated  
XX polypeptide or its fragment whose expression is inhibited by the  
XX antisense nucleic acid; (4) an antibody capable of specifically binding  
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
XX proliferation; (7) identifying a compound that influences the activity of  
XX the gene product or that has an activity against a biological pathway  
XX required for proliferation, or that inhibits cellular proliferation; (8)

identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation or to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences)

Sequence 477 AA:

Query Match 100.0%; Score 2393; DB 6; Length 477;  
Best Local Similarity 100.0%; Pred. No. 6.9e-207;  
Matches 477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MKVTLPEPERAGVWVGVDMIDRYWYGPTRISPEAPVPVYKNTIIEERPGGANVAMNI 60  
1 MKVTLPEPERAGVWVGVDMIDRYWYGPTRISPEAPVPVYKNTIIEERPGGANVAMNI 60  
61 ASIGANARLVGLTIGIDDAARALSKSLADVNVKCDPVSVPHTTTKLRLVSRNOQLRLD 120  
61 ASIGANARLVGLTIGIDDAARALSKSLADVNVKCDPVSVPHTTTKLRLVSRNOQLRLD 120  
121 FEEGFEVDPPQPLHERINQALSSIGALVLSDYAKGALASVQOMIOLARKAGVPVLIDPKG 180  
121 FEEGFEVDPPQPLHERINQALSSIGALVLSDYAKGALASVQOMIOLARKAGVPVLIDPKG 180  
181 TDFERYGATLTLPNISEFAVVGCKTEBEIYERGMKLIADVELSALVTRSEOGMSLL 240  
181 TDFERYGATLTLPNISEFAVVGCKTEBEIYERGMKLIADVELSALVTRSEOGMSLL 240  
241 QPGKAPLHPPTQAOBYVDVAGDVTIGVLAATLAAGNSIEBACFPANAAAGVVGKLG 300  
241 QPGKAPLHPPTQAOBYVDVAGDVTIGVLAATLAAGNSIEBACFPANAAAGVVGKLG 300  
301 STVSPTELENAVGRADTGFVMTBEELKLAVAARRGKGVMTNGVFDILHAGHVSYL 360  
301 STVSPTELENAVGRADTGFVMTBEELKLAVAARRGKGVMTNGVFDILHAGHVSYL 360  
361 ANARKLGDRLIVAVNSDASTKRLKGDSPVNPLEQRMIVGALBAVDWVVSFEEDTPORL 420  
361 ANARKLGDRLIVAVNSDASTKRLKGDSPVNPLEQRMIVGALBAVDWVVSFEEDTPORL 420  
421 IAGILPDLVLVKGDDYKPEEIASKEVWANGGEVLVNFEDGCSYTNIIKKIQODKKG 477  
421 IAGILPDLVLVKGDDYKPEEIASKEVWANGGEVLVNFEDGCSYTNIIKKIQODKKG 477

RESULT 5  
ADH88979 standard; protein; 477 AA.

ADH88979;

22-APR-2004 (first entry)

Escherichia coli K1 S26 mutant, SEQ ID 13.

Pathogenic microorganism; K1 S26; mutant; pathogenicity; virulence; bacterial infection; extra-intestinal infection; enterobacterium.

Escherichia coli.

Synthetic.

XX MO2004005535-A2.

XX 15-JAN-2004.

XX 09-JUL-2003; 2003MO-EP008209.

XX 09-JUL-2002; 2002EP-00008636.

XX (MUTA-) MUTABILIS SA.

XX Beacalch S;

XX WPI; 2004-091381/09.

XX N-PSDB; ADH88994.

PT Identifying and selecting a gene required for the proliferation in vivo of a pathogenic microorganism comprises determining the virulence of mutant genes on an experimental model of infection, and their effect on enteric colonization.

XX Claim 14; SEQ ID NO 13; 60pp; English.

XX The present invention relates to a method for identifying and selecting a gene required for the proliferation in vivo of a pathogenic microorganism. The method comprises determining the virulence of mutant genes on an experimental model of infection, and their effect on enteric colonization in an axenic mouse model. ADH88967-ADH88981 and ADH88982-ADH88996 are pathogenicity or virulence targets and their coding sequences which were used in the method of the invention. Compounds capable of inhibiting pathogenicity or virulence target expression are useful for developing medicinal products for inhibiting a bacterial infection, in particular an extra-intestinal infection in the case of enterobacteria.

XX Sequence 477 AA:

Query Match 100.0%; Score 2393; DB 8; Length 477;  
Best Local Similarity 100.0%; Pred. No. 6.9e-207;  
Matches 477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MKVTLPEPERAGVWVGVDMIDRYWYGPTRISPEAPVPVYKNTIIEERPGGANVAMNI 60  
1 MKVTLPEPERAGVWVGVDMIDRYWYGPTRISPEAPVPVYKNTIIEERPGGANVAMNI 60  
61 ASIGANARLVGLTIGIDDAARALSKSLADVNVKCDPVSVPHTTTKLRLVSRNOQLRLD 120  
61 ASIGANARLVGLTIGIDDAARALSKSLADVNVKCDPVSVPHTTTKLRLVSRNOQLRLD 120  
121 FEEGFEVDPPQPLHERINQALSSIGALVLSDYAKGALASVQOMIOLARKAGVPVLIDPKG 180  
121 FEEGFEVDPPQPLHERINQALSSIGALVLSDYAKGALASVQOMIOLARKAGVPVLIDPKG 180  
241 QPGKAPLHPPTQAOBYVDVAGDVTIGVLAATLAAGNSIEBACFPANAAAGVVGKLG 300  
241 QPGKAPLHPPTQAOBYVDVAGDVTIGVLAATLAAGNSIEBACFPANAAAGVVGKLG 300  
301 STVSPTELENAVGRADTGFVMTBEELKLAVAARRGKGVMTNGVFDILHAGHVSYL 360  
301 STVSPTELENAVGRADTGFVMTBEELKLAVAARRGKGVMTNGVFDILHAGHVSYL 360  
361 ANARKLGDRLIVAVNSDASTKRLKGDSPVNPLEQRMIVGALBAVDWVVSFEEDTPORL 420  
361 ANARKLGDRLIVAVNSDASTKRLKGDSPVNPLEQRMIVGALBAVDWVVSFEEDTPORL 420  
421 IAGILPDLVLVKGDDYKPEEIASKEVWANGGEVLVNFEDGCSYTNIIKKIQODKKG 477  
421 IAGILPDLVLVKGDDYKPEEIASKEVWANGGEVLVNFEDGCSYTNIIKKIQODKKG 477

RESULT 6  
 ABU28278  
 ID ABU28278 standard; protein; 476 AA.  
 AC ABU28278;  
 DT 19-JUN-2003 (first entry)  
 DE Protein encoded by prokaryotic essential gene #13805.  
 KM Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 OS Enterobacter cloacae.  
 XX MO200277183-A2.  
 XX 03-OCT-2002.  
 XX 21-MAR-2002; 2002MO-US009107.  
 XX 21-MAR-2001; 2001US-00815242.  
 XX 06-SEP-2001; 2001US-00948993.  
 XX 25-OCT-2001; 2001US-0342933P.  
 XX 08-FEB-2002; 2002US-00072851.  
 XX 06-MAR-2002; 2002US-0362699P.  
 XX (BLIT-) ELITRA PHARM INC.  
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Foreyth RA, Xu HH;  
 XX WPI, 2003-0239926/02.  
 XX N-PSDB; ACA32148.  
 DR New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX Claim 25; SEQ ID NO 56202; 1766bp; English.

SEQ Sequence 476 AA;  
 Query Match 95.9%; Score 2294; DB 6; Length 476;  
 Best Local Similarity 94.7%; Pred. No. 66-198;  
 Matches 451; Conservative 19; Mismatches 6; Indels 0; Gaps 0;

QY 1 MKVTLPEFERAGVWVGDVWLDYRWYGPTRISPEAPVPVVKVNTTEERPGAAVMMNI 60  
 DB 1 MKVTLPEFERAGVWVGDVWLDYRWYGPTRISPEAPVPVVKVNTTEERPGAAVMMNI 60  
 QY 61 ASLGANARLVGLTGIDDAARALSKSLADVNVKCDPVSVPPTITTKLRVLSRNOQLRLD 120  
 DB 61 ASLGAGSRLVGLTGIDDAARALSKSLADVNVKCDPVSVPPTITTKLRVLSRNOQLRLD 120  
 QY 121 FESGFEVDPPQPIHERINQALSSIGALVLSDYAKGALASVQOMIOLARKGVVLLDPKG 180  
 DB 121 FESGFEVDPPQPIHERINQALSSIGALVLSDYAKGALASVQOMIOLARKGVVLLDPKG 180  
 QY 121 FESGFEVDPPQPIHERINQALSSIGALVLSDYAKGALASVQOMIOLARKGVVLLDPKG 180  
 DB 121 FESGFEVDPPQPIHERINQALSSIGALVLSDYAKGALASVQOMIOLARKGVVLLDPKG 180  
 QY 181 TDPERYRGATLTPNLSPEAVVGVKCTEERIEYERGMKLADYELSLVTRSEOGMSL 240  
 DB 181 TDPERYRGATLTPNLSPEAVVGVKCTEERIEYERGMKLADYELSLVTRSEOGMSL 240  
 QY 241 QPKAPLHMPTOAQEVYDVYAGDTVYGVLAATLAAGNSLBEACFPANAAAGVVGKLG 300  
 DB 241 QPKAPLHMPTOAQEVYDVYAGDTVYGVLAATLAAGNSLBEACFPANAAAGVVGKLG 300  
 QY 301 STVSPTELENAVGRADTGFVGWTEBELKLAVAAARKGKRVMTNGVPTIIHAGHVSYL 360  
 DB 301 STVSPTELENAVGRADTGFVGWTEBELKLAVAAARKGKRVMTNGVPTIIHAGHVSYL 360  
 QY 361 ANARKGDRLLIVVNSDASTKRLKGRSPVNPPEOMIVYGALEAVDWMVSPEDPQRL 420  
 DB 361 ANARKGDRLLIVVNSDASTKRLKGRSPVNPPEOMIVYGALEAVDWMVSPEDPQRL 420  
 QY 421 IAGILPDLVKGSDYPEREISAGSKSEVWANGSEVLYNFEDGCTNNIIKKIQDDK 476  
 DB 421 IAGILPDLVKGSDYPEREISAGSKSEVWANGSEVLYNFEDGCTNNIIKKIQDDK 476

RESULT 7  
 AAU38166  
 ID AAU38166 standard; protein; 477 AA.  
 AC AAU38166;  
 DT 14-FEB-2002 (first entry)  
 DE Salmonella typhi cellular proliferation protein #57.  
 KM Antisense; prokaryotic cellular proliferation protein; antibiotic;  
 OS Salmonella typhi.  
 XX MO200170955-A2.  
 XX 27-SEP-2001.  
 XX 21-MAR-2001; 2001MO-US009180.  
 XX 21-MAR-2000; 2000US-0191078P.  
 XX 26-MAY-2000; 2000US-0206848P.  
 XX 26-MAY-2000; 2000US-0207727P.  
 XX 23-OCT-2000; 2000US-0242578P.  
 XX 27-NOV-2000; 2000US-0253625P.  
 XX 22-DEC-2000; 2000US-0257931P.  
 XX 16-FEB-2001; 2001US-0269308P.  
 XX (BLIT-) ELITRA PHARM INC.  
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto R, Xu HH;

DR WPI; 2001-611495/70.  
 DR N-PSDB; AAS56025.  
 XX  
 PT New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids.  
 XX  
 XX Example 3; SEQ ID NO 13759; 511pp; English.  
 XX  
 CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the genes,  
 CC their use in the discovery of novel antibiotics, the essential genes  
 CC themselves and the encoded proteins. The prokaryotes used are *Escherichia*  
 CC *coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*,  
 CC *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also  
 CC useful for the identification of potential new targets for antibiotic  
 CC development. The antisense nucleic acids can also be used to identify  
 CC proteins used in proliferation, to express these proteins, and to obtain  
 CC antibodies capable of binding to the expressed proteins. The proteins can  
 CC be used to screen compounds in rational drug discovery programmes. The  
 CC antisense nucleic acid sequence is also useful to screen for homologous  
 CC nucleic acids which are required for cell proliferation in a wide variety  
 CC of organisms. The present sequence represents an essential prokaryotic  
 CC cellular proliferation protein. Note: The sequence data for this patent  
 CC did not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 XX Sequence 477 AA;  
 SQ  
 Query Match 94.7%; Score 2265; DB 4; Length 477;  
 Best Local Similarity 93.5%; Pred. No. 2.5e-195;  
 Matches 445; Conservative 19; Mismatches 12; Indels 0; Gaps 0;  
 QY 1 MKVTLPEPERAGVWVGDVMDLRYWYGPTSRISPEAPVVPVYKNTIERRPGGANVAMNI 60  
 DB 1 MKVTLPEPERAGVWVGDVMDLRYWYGPTSRISPEAPVVPVYKNTIERRPGGANVAMNI 60  
 QY 61 ASLGANRLVGLTIGIDDAARALSKSLADVNYKDFVSVPHTTTKRLVLSRNQQLRLD 120  
 DB 61 ASLGANRLVGLTIGIDDAARALSKSLADVNYKDFVSVPHTTTKRLVLSRNQQLRLD 120  
 QY 121 FEEBEGVDPOPLHERINQALSSIGALVLSYAKGALASVQMIQLARKAGVPLTIPKG 180  
 DB 121 FEEBEGVDPOPLHERINQALSSIGALVLSYAKGALASVQMIQLARKAGVPLTIPKG 180  
 QY 121 FEEBEGVDPOPLHERINQALSSIGALVLSYAKGALASVQMIQLARKAGVPLTIPKG 180  
 DB 121 FEEBEGVDPOPLHERINQALSSIGALVLSYAKGALASVQMIQLARKAGVPLTIPKG 180  
 QY 181 TDFERYGATLTLPNLSFEAVVKGCKTEBEIVRGMKLIADYEALSLVLRSEQMSLT 240  
 DB 181 TDFERYGATLTLPNLSFEAVVKGCKTEBEIVRGMKLIADYEALSLVLRSEQMSLT 240  
 QY 181 TDFERYGATLTLPNLSFEAVVKGCKTEBEIVRGMKLIADYEALSLVLRSEQMSLT 240  
 DB 181 TDFERYGATLTLPNLSFEAVVKGCKTEBEIVRGMKLIADYEALSLVLRSEQMSLT 240  
 QY 241 QPKAPLHMPPTQAEVVDVDTGAGDTVIGVLAATLAAAGNSLEBACFPANAAAGVVGKLG 300  
 DB 241 QPKAPLHMPPTQAEVVDVDTGAGDTVIGVLAATLAAAGNSLEBACFPANAAAGVVGKLG 300  
 QY 241 QPKAPLHMPPTQAEVVDVDTGAGDTVIGVLAATLAAAGNSLEBACFPANAAAGVVGKLG 300  
 DB 241 QPKAPLHMPPTQAEVVDVDTGAGDTVIGVLAATLAAAGNSLEBACFPANAAAGVVGKLG 300  
 QY 301 STVPPIELNANVRGADTGFVMTBEELRQAVASARKKGEVYVNTNGVFDILAHGHSYL 360  
 DB 301 STVPPIELNANVRGADTGFVMTBEELRQAVASARKKGEVYVNTNGVFDILAHGHSYL 360  
 QY 301 STVPPIELNANVRGADTGFVMTBEELRQAVASARKKGEVYVNTNGVFDILAHGHSYL 360  
 DB 301 STVPPIELNANVRGADTGFVMTBEELRQAVASARKKGEVYVNTNGVFDILAHGHSYL 360  
 QY 361 ANARKLGDRLIVANVSDASTRLKGDSPVPLLEORMIVLGAALAVVMVVSFEEDTPORL 420  
 DB 361 ANARKLGDRLIVANVSDASTRLKGDSPVPLLEORMIVLGAALAVVMVVSFEEDTPORL 420  
 QY 421 IAGILPDLVLVKGQDYKPEELIAGSKVWANGGEVLVLFEDGCSFTNIIKKIQDDK 476  
 DB 421 IAGILPDLVLVKGQDYKPEELIAGSKVWANGGEVLVLFEDGCSFTNIIKKIQDTS 476

XX  
 DE Protein encoded by Prokaryotic essential gene #33081.  
 XX  
 XX Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 KM  
 XX Salmomella typhi.  
 OS  
 XX WO200277183-A2.  
 PN  
 XX 03-OCT-2002.  
 PD  
 XX 21-MAR-2002; 2002WO-US009107.  
 PF  
 XX 21-MAR-2001; 2001US-00815242.  
 XX 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 PR  
 XX (ELIT-) ELITRA PHARM INC.  
 PA  
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 PI WPI; 2003-029926/02.  
 DR N-PSDB; ACA51424.  
 DR  
 XX  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 PS Claim 25; SEQ ID NO 75478; 1766pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway;  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 XX Sequence 477 AA;  
 SQ  
 Query Match 94.7%; Score 2265; DB 6; Length 477;  
 Best Local Similarity 93.5%; Pred. No. 2.5e-195;  
 Matches 445; Conservative 19; Mismatches 12; Indels 0; Gaps 0;  
 QY 1 MKVTLPEPERAGVWVGDVMDLRYWYGPTSRISPEAPVVPVYKNTIERRPGGANVAMNI 60  
 DB 1 MKVTLPEPERAGVWVGDVMDLRYWYGPTSRISPEAPVVPVYKNTIERRPGGANVAMNI 60

Db 1 MKVTLPEFERAGVWVGVMDLDRYWGPTCRISPEAPVPVVKVNTVERGGANVAMNI 60  
 Qy 61 ASLGANARLVGLTGIDDAARALSKSLADVNVKCDPVSVPHTPTTKLRVLSRNOQLRLD 120  
 Db 61 ASLGANARLVGLTGIDDAARALSKSLADVNVKCDPVSVPHTPTTKLRVLSRNOQLRLD 120  
 Qy 121 FEEGFGVDPQPLHERINQALSGISGALVLSYAKGALTSVQTMISLARQGVPLIDPKG 180  
 Db 121 FEEGFGVDPQPLHERINQALSGISGALVLSYAKGALTSVQTMISLARQGVPLIDPKG 180  
 Qy 181 TDFERRYGATLTLPNISEFAVVGKCTEERIVERGKMLADVELSALVTRSEOGMSLT 240  
 Db 181 TDFERRYGATLTLPNISEFAVVGKCTEERIVERGKMLADVELSALVTRSEOGMSLT 240  
 Qy 241 QPGRAPLHPTQAOEVDVYAGDVTYIGVLAATLAAGNSLEBAACFPANAAAGVVGLGT 300  
 Db 241 QPGRAPLHPTQAOEVDVYAGDVTYIGVLAATLAAGNSLEBAACFPANAAAGVVGLGT 300  
 Qy 301 STVSPTELENAVGRADTGFVGTMBEELKLAVAARKGKGVMTNGVPDILAHGVSYL 360  
 Db 301 STVSPTELENAVGRADTGFVGTMBEELKLAVAARKGKGVMTNGVPDILAHGVSYL 360  
 Qy 361 ANARKGDRILIVANSDASTKRLKGSRPVNPLEQRMIVIGALEAVDMVVSFEEDTPQRL 420  
 Db 361 ANARKGDRILIVANSDASTKRLKGSRPVNPLEQRMIVIGALEAVDMVVSFEEDTPQRL 420  
 Qy 421 IAGILPDLVYKGGDYKPEBISKEVWANGGEVLVINFEDGCTTNIKKIQODK 476  
 Db 421 IAGILPDLVYKGGDYKPEBISKEVWANGGEVLVINFEDGCTTNIKKIQOTSE 476

## RESULT 9

ABU32353 ID ABU32353 standard; protein; 477 AA.

ABU32353;

19-JUN-2003 (first entry)

Protein encoded by prokaryotic essential gene #17880.

Antisense; prokaryotic essential gene; cell proliferation; drug design.

Klebsiella pneumoniae.

W0200277183-A2.

03-OCT-2002.

21-MAR-2002; 2002W0-US009107.

21-MAR-2001; 2001US-00815242.

25-OCT-2001; 2001US-00948993.

08-FEB-2002; 2002US-00072851.

06-MAR-2002; 2002US-0362599P.

(ELIT-) ELITRA PHARM INC.

Mang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

Wali D, Trawick JD, Carr GJ, Yamamoto R, Foreyth RA, Xu HH;

WPI, 2003-023926/02.

N-PSDB; ACA36223.

Claim 25; SEQ ID NO 60277; 1766bp; English.

The invention relates to an isolated nucleic acid comprising any one of the 613 antisense sequences given in the specification where expression

CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-regulated gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPD at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

SO Sequence 477 AA;

Query Match 93.1%; Score 2227; DB 6; Length 477;

Best Local Similarity 92.2%; Pred. No. 6; 8e-192; Matches 437; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MKVTLPEFERAGVWVGVMDLDRYWGPTCRISPEAPVPVVKVNTVERGGANVAMNI 60  
 Db 1 MKVTLPEFERAGVWVGVMDLDRYWGPTCRISPEAPVPVVKVNTVERGGANVAMNI 60  
 Qy 61 ASLGANARLVGLTGIDDAARALSKSLADVNVKCDPVSVPHTPTTKLRVLSRNOQLRLD 120  
 Db 61 ASLGANARLVGLTGIDDAARALSKSLADVNVKCDPVSVPHTPTTKLRVLSRNOQLRLD 120  
 Qy 121 FEEGFGVDPQPLHERINQALSGISGALVLSYAKGALTSVQTMISLARQGVPLIDPKG 180  
 Db 121 FEEGFGVDPQPLHERINQALSGISGALVLSYAKGALTSVQTMISLARQGVPLIDPKG 180  
 Qy 181 TDFERRYGATLTLPNISEFAVVGKCTEERIVERGKMLADVELSALVTRSEOGMSLT 240  
 Db 181 TDFERRYGATLTLPNISEFAVVGKCTEERIVERGKMLADVELSALVTRSEOGMSLT 240  
 Qy 241 QPGRAPLHPTQAOEVDVYAGDVTYIGVLAATLAAGNSLEBAACFPANAAAGVVGLGT 300  
 Db 241 QPGRAPLHPTQAOEVDVYAGDVTYIGVLAATLAAGNSLEBAACFPANAAAGVVGLGT 300  
 Qy 301 STVSPTELENAVGRADTGFVGTMBEELKLAVAARKGKGVMTNGVPDILAHGVSYL 360  
 Db 301 STVSPTELENAVGRADTGFVGTMBEELKLAVAARKGKGVMTNGVPDILAHGVSYL 360  
 Qy 361 ANARKGDRILIVANSDASTKRLKGSRPVNPLEQRMIVIGALEAVDMVVSFEEDTPQRL 420  
 Db 361 ANARKGDRILIVANSDASTKRLKGSRPVNPLEQRMIVIGALEAVDMVVSFEEDTPQRL 420  
 Qy 421 IAGILPDLVYKGGDYKPEBISKEVWANGGEVLVINFEDGCTTNIKKIQOD 474  
 Db 421 IAGILPDLVYKGGDYKPEBISKEVWANGGEVLVINFEDGCTTNIKKIQOD 474

## RESULT 10

ABO62691 ID ABO62691 standard; protein; 499 AA.

XX

AC ABO62691;  
 XX  
 DT 29-JUL-2004 (first entry)  
 XX  
 DE Klebsiella pneumoniae polypeptide seqid 9208.  
 XX  
 KW Recombinant expression vector; transcription regulatory element;  
 KM Klebsiella pneumoniae protein; antibacterial; Vaccine.  
 XX  
 OS Klebsiella pneumoniae.  
 XX  
 FN US6610836-B1.  
 XX  
 PD 26-AUG-2003.  
 XX  
 PF 27-JAN-2000; 2000US-00489039.  
 XX  
 PR 29-JAN-1999; 99US-0117747P.  
 XX  
 PA (GENO-) GENOME THERAPEUTICS CORP.  
 XX  
 PI Breton GL, Osborne M;  
 XX  
 DR WPI; 2003-895346/82.  
 DR N-PSDB; ACH96242.  
 XX  
 PT New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for  
 XX preparing a vaccine composition against Klebsiella pneumoniae.  
 PS Disclosure; SEQ ID NO 9208; 932pp; English.  
 XX  
 CC The invention describes a new isolated nucleic acid encoding a Klebsiella  
 CC pneumoniae polypeptide. Also described are: a recombinant expression  
 CC vector comprising the nucleic acid, operably linked to a transcription  
 CC regulatory element; and a cell comprising the recombinant expression  
 CC vector. The nucleic acid is useful for preparing a vaccine composition  
 CC against Klebsiella pneumoniae. This is the amino acid sequence of a  
 CC Klebsiella pneumoniae polypeptide of the invention  
 XX  
 SQ Sequence 499 AA;  
 Query Match 93.1%; Score 2227; DB 7; Length 499;  
 Best Local Similarity 92.2%; Pred. No. 7.3e-192;  
 Matches 437; Conservative 25; Mismatches 12; Indels 0; Gaps 0;  
 QY 1 MKVTLPFERAGVWVGDVMDRMYGPTSRISPEAPVYVYKVTIERPGGANVAMNI 60  
 DB 23 MKVTLPFERAGVWVGDVMDRMYGPTSRISPEAPVYVYKVTIERPGGANVAMNI 82  
 QY 61 ASLGANARLVGLTGIDDAARALSKSLADVNYKDFVSVPHTPTTKLRVLSRNOQLRLD 120  
 DB 83 ASLGATSRVLGVTGIDDAARALSKSLADVNYKDFVSVPHTPTTKLRVLSRNOQLRLD 142  
 QY 121 FEEFEGVDPPPLHERINQALSSIGALVLSYAKGALASVQOMQLARKAGVPTLIDPKG 180  
 DB 143 FEEFEGVDPPPMHRRIQALSGISGALVLSYAKGALASVQOMQLARKAGVPTLIDPKG 202  
 QY 181 TDFPRYRGATLTTPNLSEFAVWGCKTEBEIVERGMKLIADYELSLALVTRSEQMSLT 240  
 DB 203 TDFPRYRGATLTTPNLSEFAVWGCKTEBEIVERGMKLIADYELSLALVTRSEQMSLT 262  
 QY 241 QPGKAPLHPTQAOEVDVDTGAGDTVIGVLAATLAAGNSLEACFPANAAAAGVVGKLG 300  
 DB 263 QPGKAPLHPTQAOEVDVDTGAGDTVIGVLAATLAAGNSLEACFPANAAAAGVVGKLG 322  
 QY 301 STVPPELEENAVRGRADTGEVWTEBEELKAAVAARRGGEVWNTNGVFDILHAGVSYL 360  
 DB 323 STVPPELEENAVRGRADTGEVWTEBEELKAAVAARRGGEVWNTNGVFDILHAGVSYL 382  
 QY 361 ANARKLGDRLIVAVNSDASTKRLKDSRPVNPLEBOMIVLGALEAVDPVVSFEEDTQRL 420  
 DB 383 ANARKLGDRLIVAVNSDASTKRLKDSRPVNPLEBOMIVLGALEAVDPVVSFEEDTQRL 442

QY 421 IAGILPDLVKGGDYKPEBTAGSKVWANGGEVLVINFEDGCTNNIKKKIQD 474  
 DB 443 IAGILPDLVKGGDYKPEBTAGSKVWANGGEVLVINFEDGCTNNIKKKIQD 496  
 RESULT 11  
 ID ABO45244  
 AC ABO45244 standard; protein; 463 AA.  
 DT 19-JUN-2003 (first entry)  
 DE Protein encoded by prokaryotic essential gene #30771.  
 KM Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 OS Salmonella paratyphi.  
 PN WO200277183-A2.  
 PD 03-OCT-2002.  
 PF 21-MAR-2002; 2002WO-US009107.  
 PR 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX  
 PA (BLIT-) ELITRA PHARM INC.  
 XX  
 PI Wang L, Zamudio C, Malone C, Haseelbeck R, Ohlsen KU, Zyckind JW;  
 PI Wall D, Trawick JD, Carr GF, Yamamoto R, Forsyth RA, Xu HH;  
 XX  
 DR WPI; 2003-029926/02.  
 DR N-PSDB; ACA49114.  
 XX  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 PS Claim 25; SEQ ID NO 73168; 1766pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 613 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-regulated gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than S. aureus, S. typhimurium,  
 CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this



CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 463 AA;

Query Match 91.8%; Score 2196; DB 6; Length 463;

Best Local Similarity 93.5%; Pred. No. 4,1e-189; Matches 435; Conservative 18; Mismatches 10; Indels 2; Gaps 2;

QY 8 FERRAGWVVDVMDLDRYWGPTSRISPEAPVPVVKVNTIIEBPGGAANVAMNITSLGANA 67  
DB 1 FERRAGWVVDVMDLDRYWGPTSRISPEAPVPVVKVNTIIEBPGGAANVAMNITSLGANA 60  
QY 68 RLVGITGIDDAARLKSADLVNKKCPVSVPTHTTKLRVLSRNQOLRLDPBEGPEG 127  
DB 61 RLVGITGIDDAARLKSADLVNKKCPVSVPTHTTKLRVLSRNQOLRLDPBEGPEG 119  
QY 128 VDPQLHERINQALSGISGALVLSQYAKGALASVQOMIOLARKGVPVLLIDPKGDFERYR 187  
DB 120 VDPQLHERINQALSGISGALVLSQYAKGALASVQOMIOLARKGVPVLLIDPKGDFERYR 179  
QY 188 GATLLTPNLSPEFAVAVKCKTEBEIYERGMKLIDYELISALLVTRSEQMSLLQPGKAPL 247  
DB 180 GATLLTPNLSPEFAVAVKCKTEBEIYERGMKLIDYELISALLVTRSEQMSLLQPGKAPL 239  
QY 248 HMPQAOEVDYVAGGTVTGVLTAATLAAGNSLEBACFPANAAAGVVGKLGSTVSPIE 307  
DB 240 HMPQAOEVDYVAGGTVTGVLTAATLAAGNSLEBACFPANAAAGVVGKLGSTVSPIE 299  
QY 308 LBNVNRADTGFQVTEBEELKLAVAARKGKGVVMTNGFQDILHAGVSYLANARKLG 367  
DB 300 LBNVNRADTGFQVTEBEELKLAVAARKGKGVVMTNGFQDILHAGVSYLANARKLG 358  
QY 368 DRLIVANSDASTKRLKDSRPVNPLEQRMIVLGALBAVDVWVSFEEDTPQRLIAGILPD 427  
DB 359 DRLIVANSDASTKRLKDSRPVNPLEQRMIVLGALBAVDVWVSFEEDTPQRLIAGILPD 418  
QY 428 LLYKGGDYKPEBIRGSKRWANGGEVLVLFEDCCSTTNIKKIQ 472  
DB 419 LLYKGGDYKPEBIRGSKRWANGGEVLVLFEDCCSTTNIKKIQ 463

RESULT 12

ABU50241 standard; protein; 476 AA.

XX ABU50241;

DT 19-JUN-2003 (first entry)

DE Protein encoded by prokaryotic essential gene #35768.

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.

OS Versinia pectis.

XX WO200277183-A2.

PD 03-OCT-2002.

PF 21-MAR-2002; 2002MO-US009107.

PR 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362639P.

XX (ELIT-) ELITRA PHARM INC.

PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

PI Wall D, Travick JD, Carr GJ, Yamamoto R, Foreyth RA, Xu HH;

XX WPI; 2003-029926/02.

DR N-PSDB; ACAS4111.

PT New antisense nucleic acids, useful for identifying proteins or screening

PT for homologous nucleic acids required for cellular proliferation to

PT isolate candidate molecules for rational drug discovery programs.

PS Claim 25; SEQ ID NO 78165; 1766bp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

SO Sequence 476 AA;

Query Match 86.8%; Score 2076; DB 6; Length 476;

Best Local Similarity 86.1%; Pred. No. 2.9e-178; Matches 409; Conservative 32; Mismatches 34; Indels 0; Gaps 0;

QY 1 MKYTLPEFERRAGWVVDVMDLDRYWGPTSRISPEAPVPVVKVNTIIEBPGGAANVAMN 60  
DB 1 MKYTLPEFERRAGWVVDVMDLDRYWGPTSRISPEAPVPVVKVNTIIEBPGGAANVAMN 60  
QY 61 ASIGANARLVGLTGIDDAARLKSADLVNKKCPVSVPTHTTKLRVLSRNQOLRLDPBEG 120  
DB 61 ASIGANARLVGLTGIDDAARLKSADLVNKKCPVSVPTHTTKLRVLSRNQOLRLDPBEG 120  
QY 121 FEEGFGVDPQPLHERINQALSGISGALVLSQYAKGALASVQOMIOLARKGVPVLLIDPKG 180  
DB 121 FEEGFGVDPQPLHERINQALSGISGALVLSQYAKGALASVQOMIOLARKGVPVLLIDPKG 180  
QY 181 TDPERYRGATLLTPNLSPEFAVAVKCKTEBEIYERGMKLIDYELISALLVTRSEQMSLL 240  
DB 181 TDPERYRGATLLTPNLSPEFAVAVKCKTEBEIYERGMKLIDYELISALLVTRSEQMSLL 240  
QY 241 QPGKADLHMPQAOEVDYVAGGTVTGVLTAATLAAGNSLEBACFPANAAAGVVGKLGSTVSPIE 300  
DB 241 QPGKADLHMPQAOEVDYVAGGTVTGVLTAATLAAGNSLEBACFPANAAAGVVGKLGSTVSPIE 300  
QY 301 STVSPIELENAVNRGRADTGFQVTEBEELKLAVAARKGKGVVMTNGFQDILHAGVSYLANARK 360  
DB 301 STVSPIELENAVNRGRADTGFQVTEBEELKLAVAARKGKGVVMTNGFQDILHAGVSYLANARK 360  
QY 361 ANARKGDRILIVANSDASTKRLKDSRPVNPLEQRMIVLGALBAVDVWVSFEEDTPQRLIAGILPD 420

Db 361 ANARKGDRILVAVNSDASTKRLGKSPVPLEQRMVVLGALAVDWMVPFEDTQRL 420  
 Qy 421 IAGILPDLVKGDDYKPEBISKEVWANGGEVLTUNFEDGCTTNIKKIQDK 475  
 Db 421 IADILPDLVKGDDYKPEBISKEVWANGGEVLTUNFEDGCTTNIKKIQDK 475

## RESULT 13

ABU0796  
 ID ABU0796 standard; protein; 474 AA.

AC ABU0796;

DT 19-JUN-2003 (first entry)

DE Protein encoded by prokaryotic essential gene #26323.

KM Antisense; prokaryotic essential gene; cell proliferation; drug design.

OS Proteus sp.

FN W0200277183-A2.

PD 03-OCT-2002.

PF 21-MAR-2002; 2002WO-US009107.

PR 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX (ELITRA PHARM INC.

PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

PI Wall D, Travick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI: 2003-029926/02.

DR N-PSDB; ACRA44666.

PT New antisense nucleic acids, useful for identifying proteins or screening

PT for homologous nucleic acids required for cellular proliferation to

PT isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 68720; 1766bp; English.

CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids

CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIFO at  
 CC ftp.wifo.int/pub/published\_pct\_sequences

XX Sequence 474 AA;

Query Match 83.9%; Score 2007; DB 6; Length 474;

Best Local Similarity 82.2%; Pred. No. 5e-172; Matches 389; Conservative 43; Mismatches 41; Indels 0; Gaps 0;

Qy 1 MKVTLPEFEPAGVWVGVNDVLDKRYWGPSTRSISPEAPVPVVKVNTIEERGGANVAMNI 60  
 Db 1 MKVTLPEFEPAGVWVGVNDVLDKRYWGPSTRSISPEAPVPVVKVNTIEERGGANVAMNI 60  
 Qy 61 ASIGANARLVGLTGIDDAARALSKSLADVNVKCDPVSPVPTHTYTKRLVLSRNOQLRLD 120  
 Db 61 ASIGANARLVGLTGIDDAARALSKSLADVNVKCDPVSPVPTHTYTKRLVLSRNOQLRLD 120  
 Qy 121 FEERGEGVDPQPIHERINQALSSIGALVSDVAKGALASVQOMIOLARKVGPVLIDPKG 180  
 Db 121 FEERGEGVDPQPIHERINQALSSIGALVSDVAKGALASVQOMIOLARKVGPVLIDPKG 180  
 Qy 181 TDFERYRGATLTPNLSPEFAVVGKCKTEBEIYERGMKLIADVELSALLVTRSEOGMSLL 240  
 Db 181 TDFERYRGATLTPNLSPEFAVVGKCKTEBEIYERGMKLIADVELSALLVTRSEOGMSLL 240  
 Qy 241 QPKAPLHMFQAOBYVDYTAGDTYIVGLAATLAAAGNSLSERACFPANAAAGVVGKLTGT 300  
 Db 241 QPKAPLHMFQAOBYVDYTAGDTYIVGLAATLAAAGNSLSERACFPANAAAGVVGKLTGT 300  
 Qy 301 STVSPTELENAVAVGRADTGVWTEBEELKLAVAARKRGKVMYMTGVPIILAHGVSYL 360  
 Db 301 STVSPTELENAVAVGRADTGVWTEBEELKLAVAARKRGKVMYMTGVPIILAHGVSYL 360  
 Qy 361 ANARKGDRILVAVNSDASTKRLKGSRPVNPLEQRMVVLGALAVDWMVPFEDTQRL 420  
 Db 361 ANARKGDRILVAVNSDASTKRLKGSRPVNPLEQRMVVLGALAVDWMVPFEDTQRL 420  
 Qy 421 IAGILPDLVKGDDYKPEBISKEVWANGGEVLTUNFEDGCTTNIKKIQ 475  
 Db 421 IAGILPDLVKGDDYKPEBISKEVWANGGEVLTUNFEDGCTTNIKKIQ 475

## RESULT 14

ADP06349  
 ID ADP06349 standard; protein; 491 AA.

AC ADP06349;

DT 12-FEB-2004 (first entry)

DE Bacterial polypeptide #2462.

XX Proteus mirabilis infection; bacterial infection; antibacterial;

KM immunostimulant.

OS Proteus mirabilis.

FN US6605709-B1.

PD 12-AUG-2003.

PF 05-APR-2000; 2000US-00543681.

PR 09-APR-1999; 99US-0128706P.

PA (GENO-) GENOME THERAPEUTICS CORP.

PI Breton GU;

DR WPI, 2003-895291/82.  
 DR N-PSDB; ADF02177.  
 XX New Proteus mirabilis polypeptides and polynucleotides, useful as  
 PT reagents for diagnosis of bacterial disease, as components of  
 PT antibacterial vaccines, as targets for antibacterial drugs, or as  
 PT biocontrol agents for plants.  
 XX  
 PS Disclosure; SEQ ID NO 6634; 870bp; English.  
 XX  
 CC The invention relates to new Proteus mirabilis polypeptides and  
 CC polynucleotides. The invention also relates to antibodies against the  
 CC polypeptides, methods for producing the polypeptides, a method of  
 CC generating vaccines for immunising an individual against P. mirabilis, a  
 CC method for evaluating a compound for the ability to bind a P. mirabilis  
 CC polypeptide and a method for screening test compounds for anti-bacterial  
 CC activity. The polypeptides and polynucleotides are useful as molecular  
 CC targets for diagnosis, preventing and treating pathological conditions  
 CC resulting from bacterial infection, as reagents for diagnosis of  
 CC bacterial diseases, as components of antibacterial vaccines, as targets  
 CC for antibacterial drugs or as bio-control agents for plants. This  
 CC sequence represents a Proteus mirabilis polypeptide of the invention.  
 CC  
 XX Sequence 491 AA;  
 SQ  
 Query Match 83.9%; Score 2007; DB 7; Length 491;  
 Best Local Similarity 82.2%; Pred. No. 5.2e-172;  
 Matches 389; Conservative 43; Mismatches 41; Indels 0; Gaps 0;  
 QY 1 MKVTLPEERAGVWVGDVMDLDRVWYGPTRISPEAPVYVNTIERRPGANVAMNI 60  
 DB 18 MKVTLPEERAGVWVGDVMDLDRVWYGPTRISPEAPVYVNTIERRPGANVAMNI 77  
 QY 61 ASIGANARLVGLTGIDDAARALSKSLADVNYKCDPVSVPTHTTKLRVLSRNOQLRLD 120  
 DB 78 ASIGANARLVGLTGIDDAARALSKSLADVNYKCDPVSVPTHTTKLRVLSRNOQLRLD 137  
 QY 121 PEEFEGVDPPPLHERINQALSSIGALVLSYAKGALASVOQMIQLARKGVPVLIDPKG 180  
 DB 138 PEEFEGVDPPPLHERINQALSSIGALVLSYAKGALASVOQMIQLARKGVPVLIDPKG 197  
 QY 181 TDFERYGATLTPNLSEFAVVGKCTEERIYERGMKLIADYELSLVTRSEOGMSL 240  
 DB 198 GDFERYGATLTPNLSEFAVVGKCTEERIYERGMKLIADYELSLVTRSEOGMSL 257  
 QY 241 QPGKAPLHMPPTQAEVVDVDTGAGDTVIGVLAATLAAGNSLEBACFPANAAAGVVGKLTG 300  
 DB 258 QANBAPLHLPQAEVVDVDTGAGDTVIGVLAATLAAGNSLEBACFPANAAAGVVGKLTG 317  
 QY 301 STVSPFIELENNVRGADTGFVMTTEBEIKLAVAAARKGKGVNTNGVFDILAHGVSYL 360  
 DB 318 STVSPFIELENNVRGADTGFVMTTEBEIKLAVAAARKGKGVNTNGVFDILAHGVSYL 377  
 QY 361 ANARKLGDRLIVANSDASTRLKGDSPVPLBORMIVGALBAVMMVVFEEEDTFORL 420  
 DB 378 ANARKLGDRLIVANSDASTRLKGDSPVPLBORMIVGALBAVMMVVFEEEDTFORL 437  
 QY 421 IAGILPDLVKGADYKPEELIAGSKRWANGGEVLVNEEDCGSTNNIIKKIQ 473  
 DB 438 IASVLPDLVKGADYKPEELIAGSKRWANGGEVLVNEEDCGSTNNIIKKIK 490  
 RESULT 15  
 ABM67735  
 ID ABM67735 standard; protein; 475 AA.  
 XX  
 AC ABM67735;  
 XX  
 DT 20-NOV-2003 (first entry)  
 XX  
 DE Photorhabdus luminescens protein sequence #832.  
 XX  
 KM Antibacterial, fungicide, insecticide, polymorphism; genetic analysis;

KM detection; food; gene expression; plant; animal; microorganism; toxin;  
 KM antibiotic; biopesticide; virulence factor; disease model; plague;  
 KM whooping cough.  
 XX  
 XX Photorhabdus luminescens.  
 OS  
 XX WO200294867-A2.  
 PN  
 XX 28-NOV-2002.  
 PD  
 XX  
 XX 07-FEB-2002; 2002MO-IB003040.  
 PF  
 XX 07-FEB-2001; 2001FR-00001659.  
 PR  
 XX (INSP) INST PASTER.  
 PA (CNRS) CNRS CENT NAT RECH SCI.  
 PI Duchaud B, Taourit S, Glaeser P, Frangeul L, Kunat F, Danchin A,  
 PI Buchrieser C;  
 XX  
 DR WPI, 2003-148459/14.  
 DR  
 XX  
 PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides,  
 PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.  
 XX  
 XX Claim 2; SEQ ID NO 832; 1205bp; French.  
 PS  
 CC The invention relates to the isolation of genes and their encoded  
 CC proteins from Photorhabdus luminescens. The isolated sequences are  
 CC sources of probes and primers for detecting the genome of P. luminescens  
 CC and related species; to study polymorphisms; for gene analysis and for  
 CC detection/amplification of the genes. Antibodies (Ab) raised against the  
 CC polypeptides encoded by the genes are used for detection/identification  
 CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that  
 CC carry a gene-containing vector are used to select compounds that  
 CC modulate, regulate, induce or inhibit expression of the genes in plants,  
 CC animals or microorganisms other than P. luminescens and are able to alter  
 CC response or sensitivity to toxins and antibiotics produced by P.  
 CC luminescens. Cells transformed to express the genes are useful for  
 CC recombinant production of the proteins, particularly toxins and  
 CC antibacterials useful as insecticides, bactericides and fungicides. The  
 CC genes, proteins, vectors containing the genes and Ab are also useful  
 CC therapeutically to treat microbial infection by bacteria or fungi that  
 CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as  
 CC biopesticides. Other uses of the genes and the proteins are as virulence  
 CC factors and for identifying targets of human diseases for which P.  
 CC luminescens is a model (particularly plague and whooping cough). This  
 CC sequence represents one of the isolated P. luminescens proteins  
 CC  
 XX  
 SQ Sequence 475 AA;  
 Query Match 81.5%; Score 1950; DB 6; Length 475;  
 Best Local Similarity 80.6%; Pred. No. 7e-167;  
 Matches 389; Conservative 36; Mismatches 54; Indels 0; Gaps 0;  
 QY 1 MKVTLPEERAGVWVGDVMDLDRVWYGPTRISPEAPVYVNTIERRPGANVAMNI 60  
 DB 1 MKVTLPEERAGVWVGDVMDLDRVWYGPTRISPEAPVYVNTIERRPGANVAMNI 60  
 QY 61 ASIGANARLVGLTGIDDAARALSKSLADVNYKCDPVSVPTHTTKLRVLSRNOQLRLD 120  
 DB 61 ASIGANARLVGLTGIDDAARALSKSLADVNYKCDPVSVPTHTTKLRVLSRNOQLRLD 120  
 QY 121 PEEFEGVDPPPLHERINQALSSIGALVLSYAKGALASVOQMIQLARKGVPVLIDPKG 180  
 DB 121 PEEFEGVDPPPLHERINQALSSIGALVLSYAKGALASVOQMIQLARKGVPVLIDPKG 180  
 QY 181 TDFERYGATLTPNLSEFAVVGKCTEERIYERGMKLIADYELSLVTRSEOGMSL 240  
 DB 181 TDFERYGATLTPNLSEFAVVGKCTEERIYERGMKLIADYELSLVTRSEOGMSL 240  
 QY 241 QPGKAPLHMPPTQAEVVDVDTGAGDTVIGVLAATLAAGNSLEBACFPANAAAGVVGKLTG 300  
 DB 241 QPGKAPLHMPPTQAEVVDVDTGAGDTVIGVLAATLAAGNSLEBACFPANAAAGVVGKLTG 300

Db	241	SVDPPLHLPTQAOBVEFDVTGAGDVTIGVLTATAIAGKPLNEACFLANAAAGVVGKLG	300
Qy	301	STVSPLELENVNRGRADTSGVMTBEEIKLVAARKGKGVMTNGVEDILHAGHVS	360
Db	301	STVSPLELENVNRGRADTSGVMTBEEIKLVAARKGKGVMTNGVEDILHAGHVS	360
Qy	361	ANARKLGDRLLVAVNSDASTRLKGDSPVNPLEQRMIVLGALBAVDVWVSFEEDTPQRL	420
Db	361	ENARKLGDRLLVAVNSDASTRLKGDSPVNPLEQRMIVLGALBAVDVWVSFEEDTPQRL	420
Qy	421	IAGILPDLVLVKGSDYKPEELNGSKVWANGGEVLVNFEDGCTTNIKKIQDDK	475
Db	421	IADVLPDVLVKGSDYKPEELNGSKVWANGGEVLVNFEDGCTTNIKKIQDDK	475

Search completed: March 18, 2006, 06:02:33  
 Job time : 193 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

# OM protein - protein search, using sw model

Run on: March 18, 2006, 06:02:48, Search time 40 Seconds  
(without alignments)  
1147.385 Million cell updates/sec

Title: US-10-520-820-13  
Perfect score: 2393  
Sequence: 1 MKVTLPEFERAGVWVGDVW.....FEDGCTWIKIKIQDDKKG 477

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database:

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2393	100.0	477	1 B65093	ADP-heptose synthase
2	2385	99.7	477	2 G91120	ADP-heptose synthase
3	2385	99.7	477	2 F85965	probable DP-heptos
4	2265	94.7	477	2 AG0891	ADP-heptose synthase
5	2076	86.8	476	2 A10080	ADP-heptose synthase
6	1712	71.5	476	1 C64127	ADP-heptose synthase
7	1350.5	56.4	474	2 A83022	LPS biosynthesis P
8	904.5	37.8	483	2 F87700	rfag protein [impo
9	863	36.1	461	2 H81319	probable ADP-hepto
10	819.5	34.2	463	2 C71887	probable adp-d-gly
11	811	33.9	315	2 G84936	kinase [imported]
12	809.5	33.8	461	1 B64627	ADP-heptose synthase
13	753.5	31.5	323	2 H81154	ADP-heptose synthase
14	745.5	31.2	323	2 H81951	probable DP-heptos
15	594	24.8	463	2 T34841	probable bifunctio
16	564.5	23.6	315	2 H70350	ADP-heptose synthase
17	391.5	16.4	157	2 H70317	glycerol-3-phospha
18	348.5	14.6	168	2 C81009	aut protein NME207
19	340.5	14.2	168	2 G82031	probable transfe
20	311.5	13.0	164	2 T39548	aut protein - Aica
21	201.5	8.4	319	2 A83882	fructokinase BH185
22	197	8.2	308	2 B83403	fructokinase PA1950
23	194	8.1	299	2 E72311	fructokinase - Therm
24	192.5	8.0	298	2 AH0947	probable sugar kin
25	189.5	7.9	148	2 A75041	probable glycerol
26	188	7.9	307	2 AB2029	ribokinase [import
27	184.5	7.7	293	1 D69690	ribokinase (EC 2.7
28	184.5	7.7	303	2 B87367	ribokinase [import
29	183.5	7.7	319	2 AC0942	probable carbohydr

30	181.5	7.6	148	2 H71120	hypothetical prote
31	181.5	7.6	378	2 F86307	hypothetical prote
32	179.5	7.5	383	2 D75288	carbohydrate kinas
33	179	7.5	137	2 A69427	glycerol-3-phospha
34	178	7.4	306	2 B64073	ribokinase (EC 2.7
35	176	7.4	294	2 H84115	ribokinase rbsK [l
36	175	7.3	311	2 D75285	carbohydrate kinas
37	174	7.3	303	2 F84401	hypothetical prote
38	174	7.3	316	2 A96852	fructokinase [impo
39	173.5	7.3	300	2 F91229	fructokinase [l
40	173.5	7.3	300	2 E86076	probable kinase y1
41	170.5	7.1	305	2 T44955	ribokinase (EC 2.7
42	170.5	7.1	343	2 D75260	probable carbohydr
43	170	7.1	300	2 A75599	ribokinase - Deino
44	169	7.1	309	1 K1E3RB	ribokinase (EC 2.7
45	169	7.1	309	2 F91215	ribokinase [import

## ALIGNMENTS

### RESULT 1

B65093

ADP-heptose synthase homolog - *Escherichia coli* (strain K-12)

N:Alternate names: hypothetical protein b3052

C:Species: *Escherichia coli*

C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004

C/Accession: B65093

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Coi

.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A>Title: The complete genome sequence of *Escherichia coli* K-12.

A/Reference number: A64720; M01D:9742617; EMD:9278503

A/Accession: B65093

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-477 <BLAT>

A/Cross-references: UNIPROT:P76558; UNIPARC:UPI000003EB75; GB:AB000387; GB:U00096; NID:93

A/Experimental source: strain K-12, substrain MG1655

C/Superfamily: hypothetical protein b3052

Query Match	100.0%	Score 2393;	DB 1;	Length 477;
Best local similarity	100.0%	Pred. No. 5.4e-143;		
Matches	477;	Conservative	0;	Mismatches 0;
		Indels	0;	Gaps 0;
QY	1	MKVTLPEFERAGVWVGDVWLDRIWYGPTSRISPEAPVVKVNTTEERPGGAANYAMNI	60	
DB	1	MKVTLPEFERAGVWVGDVWLDRIWYGPTSRISPEAPVVKVNTTEERPGGAANYAMNI	60	
QY	61	ASIGANARLVGLTGIDDAARLSKSLADVNVKDPVSVPHTPTTLRLVSRNQOLIRLD	120	
DB	61	ASIGANARLVGLTGIDDAARLSKSLADVNVKDPVSVPHTPTTLRLVSRNQOLIRLD	120	
QY	121	FEFGFEGVDPQPLHERINQALSSIGALVLSDYAKGALASVOQMIQLARKAGVFLIDPKG	180	
DB	121	FEFGFEGVDPQPLHERINQALSSIGALVLSDYAKGALASVOQMIQLARKAGVFLIDPKG	180	
QY	181	TDFERYRGATLLTPNLSPEPAVVGKCTEERIEVERGKLIADYELGALLVTRSEQMSLL	240	
DB	181	TDFERYRGATLLTPNLSPEPAVVGKCTEERIEVERGKLIADYELGALLVTRSEQMSLL	240	
QY	241	QPKAPLHPPTQOQVYVDVGTADTVYIGVLAATLAAGNSLEBAQCFPANAAGVVGKLTG	300	
DB	241	QPKAPLHPPTQOQVYVDVGTADTVYIGVLAATLAAGNSLEBAQCFPANAAGVVGKLTG	300	
QY	301	STSPFIELENAVGRADTQGVWTEBELKLAVAARKEGKVMVTGVPDIILAGVSYL	360	
DB	301	STSPFIELENAVGRADTQGVWTEBELKLAVAARKEGKVMVTGVPDIILAGVSYL	360	
QY	361	ANARKIGDRLIIVAVNSDASTKRLKSDRPVNPLEQPMIVIGALBAVDVWVSFEEDTPORL	420	
DB	361	ANARKIGDRLIIVAVNSDASTKRLKSDRPVNPLEQPMIVIGALBAVDVWVSFEEDTPORL	420	

QY 421 IAGILPDLVKGDDYKPEBIAGSKVWANGGEVLVLPFBDCSTTNIKKIQDDKKG 477  
 |||||  
 DB 421 IAGILPDLVKGDDYKPEBIAGSKVWANGGEVLVLPFBDCSTTNIKKIQDDKKG 477

## RESULT 2

G91120  
 ADP-heptose synthase [imported] - Escherichia coli (strain O157:H7, substrain RIMD 05099)  
 C:Species: Escherichia coli  
 C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004  
 C/Accession: G91120  
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;  
 gasawara, N.; Yasunaga, T.; Kuhnara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Ref: 8, 11-22, 2001  
 A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene  
 A:Reference number: A96629; MUID:21156231; PMID:11258796  
 A/Accession: G91120  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-477 <HAV>  
 A/Cross-references: UNIPROT:Q8XBW4; UNIPARC:UPI000000591; GB:BA000007; PIDN:BA037358.1;  
 A:Experimental source: strain O157:H7, substrain RIMD 0509952  
 C/Genetics:  
 A:Gene: ECa3935  
 C/Superfamily: hypothetical protein b3052

Query Match 99.7%; Score 2385; DB 2; Length 477;  
 Best Local Similarity 99.6%; Pred. No. 1.7e-142;  
 Matches 475; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKVTLPEFERAGVWVGVMDLDRYWGPTSRISPEAPVPVKKVTIEERPGAAVAMNI 60  
 |||||  
 DB 1 MKVTLPEFERAGVWVGVMDLDRYWGPTSRISPEAPVPVKKVTIEERPGAAVAMNI 60  
 |||||  
 QY 61 ASIGANARLVGLTGIDDAARALSKSLADVNVKCDPVSPVPTITTKLRVLSRNOQLRLD 120  
 |||||  
 DB 61 ASIGANARLVGLTGIDDAARALSKSLADVNVKCDPVSPVPTITTKLRVLSRNOQLRLD 120  
 |||||  
 QY 121 FEEGFEVDPOPLHERINQALSSIGALVLSDYAKGALASVOQMIQLARKAGVPLIDPKG 180  
 |||||  
 DB 121 FEEGFEVDPOPLHERINQALSSIGALVLSDYAKGALASVOQMIQLARKAGVPLIDPKG 180  
 |||||  
 QY 181 TDFERYGATLTPNLSEFAVVGKCTEIEVERGKMLADVELSALLVTRSEQMSLL 240  
 |||||  
 DB 181 TDFERYGATLTPNLSEFAVVGKCTEIEVERGKMLADVELSALLVTRSEQMSLL 240  
 |||||  
 QY 241 QPGKAPLHMPQAOEYVDYTGAGDTVIGVLAATLAAGNSLEBACFPANAAAGVVGKLG 300  
 |||||  
 DB 241 QPGKAPLHMPQAOEYVDYTGAGDTVIGVLAATLAAGNSLEBACFPANAAAGVVGKLG 300  
 |||||  
 QY 301 STVSPIELENAVRGRADTGFVMTSEBELKLAVAARKRGKRVMTNGVFDILHAGVSYL 360  
 |||||  
 DB 301 STVSPIELENAVRGRADTGFVMTSEBELKLAVAARKRGKRVMTNGVFDILHAGVSYL 360  
 |||||  
 QY 361 ANARKGDRILIVANSDASTKRLKGSRPVPLEQRMIVIGALBAVDMVVSFEEDTPQRL 420  
 |||||  
 DB 361 ANARKGDRILIVANSDASTKRLKGSRPVPLEQRMIVIGALBAVDMVVSFEEDTPQRL 420  
 |||||  
 QY 421 IAGILPDLVKGDDYKPEBIAGSKVWANGGEVLVLPFBDCSTTNIKKIQDDKKG 477  
 |||||  
 DB 421 IAGILPDLVKGDDYKPEBIAGSKVWANGGEVLVLPFBDCSTTNIKKIQDDKKG 477

## RESULT 3

F85965  
 Probable kinase Z4405 [imported] - Escherichia coli (strain O157:H7, substrain EDL933)  
 C:Species: Escherichia coli  
 C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004  
 C/Accession: F85965  
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
 Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potemoulsis, K.; Apodaca,  
 Nature 409, 529-533, 2001  
 A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A/Accession: F85965  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-477 <STO>  
 A/Cross-references: UNIPROT:Q8XBW4; UNIPARC:UPI000000591; GB:AE005174; NID:G12517628; P  
 A:Experimental source: strain O157:H7, substrain EDL933  
 C/Genetics:  
 A:Gene: Z4405  
 C/Superfamily: hypothetical protein b3052

Query Match 99.7%; Score 2385; DB 2; Length 477;  
 Best Local Similarity 99.6%; Pred. No. 1.7e-142;  
 Matches 475; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKVTLPEFERAGVWVGVMDLDRYWGPTSRISPEAPVPVKKVTIEERPGAAVAMNI 60  
 |||||  
 DB 1 MKVTLPEFERAGVWVGVMDLDRYWGPTSRISPEAPVPVKKVTIEERPGAAVAMNI 60  
 |||||  
 QY 61 ASIGANARLVGLTGIDDAARALSKSLADVNVKCDPVSPVPTITTKLRVLSRNOQLRLD 120  
 |||||  
 DB 61 ASIGANARLVGLTGIDDAARALSKSLADVNVKCDPVSPVPTITTKLRVLSRNOQLRLD 120  
 |||||  
 QY 121 FEEGFEVDPOPLHERINQALSSIGALVLSDYAKGALASVOQMIQLARKAGVPLIDPKG 180  
 |||||  
 DB 121 FEEGFEVDPOPLHERINQALSSIGALVLSDYAKGALASVOQMIQLARKAGVPLIDPKG 180  
 |||||  
 QY 181 TDFERYGATLTPNLSEFAVVGKCTEIEVERGKMLADVELSALLVTRSEQMSLL 240  
 |||||  
 DB 181 TDFERYGATLTPNLSEFAVVGKCTEIEVERGKMLADVELSALLVTRSEQMSLL 240  
 |||||  
 QY 241 QPGKAPLHMPQAOEYVDYTGAGDTVIGVLAATLAAGNSLEBACFPANAAAGVVGKLG 300  
 |||||  
 DB 241 QPGKAPLHMPQAOEYVDYTGAGDTVIGVLAATLAAGNSLEBACFPANAAAGVVGKLG 300  
 |||||  
 QY 301 STVSPIELENAVRGRADTGFVMTSEBELKLAVAARKRGKRVMTNGVFDILHAGVSYL 360  
 |||||  
 DB 301 STVSPIELENAVRGRADTGFVMTSEBELKLAVAARKRGKRVMTNGVFDILHAGVSYL 360  
 |||||  
 QY 361 ANARKGDRILIVANSDASTKRLKGSRPVPLEQRMIVIGALBAVDMVVSFEEDTPQRL 420  
 |||||  
 DB 361 ANARKGDRILIVANSDASTKRLKGSRPVPLEQRMIVIGALBAVDMVVSFEEDTPQRL 420  
 |||||  
 QY 421 IAGILPDLVKGDDYKPEBIAGSKVWANGGEVLVLPFBDCSTTNIKKIQDDKKG 477  
 |||||  
 DB 421 IAGILPDLVKGDDYKPEBIAGSKVWANGGEVLVLPFBDCSTTNIKKIQDDKKG 477

## RESULT 4

AG0891  
 ADP-heptose synthase (EC 2.7.-.-) [imported] - Salmonella enterica subsp. enterica serov.  
 A:Species: Salmonella enterica subsp. enterica serovar Typh  
 A>Note: this species has also been called Salmonella typhi  
 C>Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
 C/Accession: AG0891  
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
 th, T.; Conerton, P.; Crofton, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
 S.; Moule, S.; O'Garra, P.  
 Nature 413, 848-852, 2001  
 A/Authors: Parky, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
 A>Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
 A:Reference number: AB0502; MUID:21534947; PMID:11677608  
 A/Accession: AG0891  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-477 <PAR>  
 A/Cross-references: UNIPARC:UPI000005A4A3; GB:AL513382; PIDN:CAD07725.1; PID:G16504277; C  
 C/Genetics:  
 A:Gene: rfaB  
 C/Superfamily: hypothetical protein b3052  
 C/Keywords: phosphotransferase

Query Match 94.7%; Score 2265; DB 2; Length 477;

Best Local Similarity 93.5%; Pred. No. 66-135;  
Matches 445; Conservative 19; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKVTLPEERAGVAVGVMDLDRYWGPTSHISEAPVYVKNVTIERPGGAANVANI 60  
DB 1 MKVTLPEERAGVAVGVMDLDRYWGPTSHISEAPVYVKNVTIERPGGAANVANI 60  
QY 61 ASLGANRLVGLTIGIDDAARALSKSLADVNVKCDPVSPHPTITKLRVSRNOQLRLD 120  
DB 61 ASLGANRLVGLTIGIDDAARALSKSLADVNVKCDPVSPHPTITKLRVSRNOQLRLD 120  
QY 121 FEESFEGVDPPQLHERINQALSSIGALVLSDYAKGALASVOOMIQLARKAGVPLIDPKG 180  
DB 121 FEESFEGVDPPQLHERINQALSSIGALVLSDYAKGALASVOOMIQLARKAGVPLIDPKG 180  
QY 181 TDFERYRGATLTLPNLSFEFAVAVGCKTEBEIYERGMKLIADYELASLVTRSEOGMTLL 240  
DB 181 TDFERYRGATLTLPNLSFEFAVAVGCKTEBEIYERGMKLIADYELASLVTRSEOGMTLL 240  
QY 241 QPGKAPLHMPTOAOEVVDYAGDVTGVLAAATLAAGNSLEBACFPANAAAGVVGKLTGT 300  
DB 241 QPGKAPLHMPTOAOEVVDYAGDVTGVLAAATLAAGNSLEBACFPANAAAGVVGKLTGT 300  
QY 301 STVSPLEENAVRGADTGFVMTBEBELKLAVAARKGKRVVNTGVFDILHAGHVSYL 360  
DB 301 STVSPLEENAVRGADTGFVMTBEBELKLAVAARKGKRVVNTGVFDILHAGHVSYL 360  
QY 361 ANARKDRLIVAVNSDASTRLKGSRPVNPLEFORMITVGALEAVDVMVVSFEEDTPORL 420  
DB 361 ANARKDRLIVAVNSDASTRLKGSRPVNPLEFORMITVGALEAVDVMVVSFEEDTPORL 420  
QY 421 IAGILPDLVKGGDYKPEEIRAGSKVWANGSEVVLNFEDECSSTNIIKKIQODK 476  
DB 421 IAGILPDLVKGGDYKPEEIRAGSKVWANGSEVVLNFEDECSSTNIIKKIQODK 476

## RESULT 5

ADP-heptose synthase (EC 2.7.-.-) [imported] - Yersinia pestis (strain CO92)  
C/Species: Yersinia pestis  
C/Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Jul-2004  
C/Accession: A10080  
R/Parikh, J.; Wren, B.W.; Thomson, N.R.; Tibbali, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; H, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett, Nature 413, 523-527, 2001  
A/Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A/Reference number: AB0001; MUID:21470413; PMID:11586360  
A/Accession: A10080  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-476 <TR>  
A/Cross-references: UNIPROT:Q82160; UNIPARC:UPI00000CD706; GB:AL590842; PIDN:CAC89508.1;  
C/Genetics:  
A/Genes: rfaB  
C/Superfamily: hypothetical protein b3052  
C/Keywords: phosphotransferase

Query Match 86.8%; Score 2076; DB 2; Length 476;  
Best Local Similarity 86.1%; Pred. No. 4.66-123;  
Matches 409; Conservative 32; Mismatches 34; Indels 0; Gaps 0;

QY 1 MKVTLPEERAGVAVGVMDLDRYWGPTSHISEAPVYVKNVTIERPGGAANVANI 60  
DB 1 MKVTLPEERAGVAVGVMDLDRYWGPTSHISEAPVYVKNVTIERPGGAANVANI 60  
QY 61 ASLGANRLVGLTIGIDDAARALSKSLADVNVKCDPVSPHPTITKLRVSRNOQLRLD 120  
DB 61 ASLGANRLVGLTIGIDDAARALSKSLADVNVKCDPVSPHPTITKLRVSRNOQLRLD 120  
QY 121 FEESFEGVDPPQLHERINQALSSIGALVLSDYAKGALASVOOMIQLARKAGVPLIDPKG 180  
DB 121 FEESFEGVDPPQLHERINQALSSIGALVLSDYAKGALASVOOMIQLARKAGVPLIDPKG 180

QY 181 TDFERYRGATLTLPNLSFEFAVAVGCKTEBEIYERGMKLIADYELASLVTRSEOGMTLL 240  
DB 181 TDFERYRGATLTLPNLSFEFAVAVGCKTEBEIYERGMKLIADYELASLVTRSEOGMTLL 240

QY 241 QPGKAPLHMPTOAOEVVDYAGDVTGVLAAATLAAGNSLEBACFPANAAAGVVGKLTGT 300  
DB 241 QPGKAPLHMPTOAOEVVDYAGDVTGVLAAATLAAGNSLEBACFPANAAAGVVGKLTGT 300

QY 301 STVSPLEENAVRGADTGFVMTBEBELKLAVAARKGKRVVNTGVFDILHAGHVSYL 360  
DB 301 STVSPLEENAVRGADTGFVMTBEBELKLAVAARKGKRVVNTGVFDILHAGHVSYL 360

QY 361 ANARKDRLIVAVNSDASTRLKGSRPVNPLEFORMITVGALEAVDVMVVSFEEDTPORL 420  
DB 361 ANARKDRLIVAVNSDASTRLKGSRPVNPLEFORMITVGALEAVDVMVVSFEEDTPORL 420

QY 421 IAGILPDLVKGGDYKPEEIRAGSKVWANGSEVVLNFEDECSSTNIIKKIQODK 476  
DB 421 IAGILPDLVKGGDYKPEEIRAGSKVWANGSEVVLNFEDECSSTNIIKKIQODK 476

## RESULT 6

ADP-heptose synthase homolog - Haemophilus influenzae (strain Rd KW20)  
C/Species: Haemophilus influenzae  
C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004  
C/Accession: C64127  
R/Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Geoghegan, N.S.M.; Science 269, 496-512, 1995  
A/Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Frazer, C.M.; Smith, H.O.; Venter, A./Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A/Reference number: A64000; MUID:95350630; PMID:7542800  
A/Accession: C64127  
A/Status: nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-476 <TR>  
A/Cross-references: UNIPROT:O05074; UNIPARC:UPI00001337AB; GB:U32828; GB:L42023; NID:g15;  
C/Superfamily: hypothetical protein b3052

Query Match 71.5%; Score 1712; DB 1; Length 476;  
Best Local Similarity 71.0%; Pred. No. 3.56-100;  
Matches 331; Conservative 56; Mismatches 77; Indels 0; Gaps 0;

QY 7 EFRAGVAVGVMDLDRYWGPTSHISEAPVYVKNVTIERPGGAANVANI 66  
DB 7 EFRAGVAVGVMDLDRYWGPTSHISEAPVYVKNVTIERPGGAANVANI 66  
QY 67 ARVGLTIGIDDAARALSKSLADVNVKCDPVSPHPTITKLRVSRNOQLRLD 126  
DB 67 ARVGLTIGIDDAARALSKSLADVNVKCDPVSPHPTITKLRVSRNOQLRLD 126  
QY 127 GVPQPLHERINQALSSIGALVLSDYAKGALASVOOMIQLARKAGVPLIDPKGTDERY 186  
DB 127 GVPQPLHERINQALSSIGALVLSDYAKGALASVOOMIQLARKAGVPLIDPKGTDERY 186  
QY 187 RGATLTLPNLSFEFAVAVGCKTEBEIYERGMKLIADYELASLVTRSEOGMTLL 246  
DB 187 RGATLTLPNLSFEFAVAVGCKTEBEIYERGMKLIADYELASLVTRSEOGMTLL 246  
QY 247 LHMPTAOEVVDYAGDVTGVLAAATLAAGNSLEBACFPANAAAGVVGKLTGTSTVS 306  
DB 247 LHMPTAOEVVDYAGDVTGVLAAATLAAGNSLEBACFPANAAAGVVGKLTGTSTVS 306  
QY 307 ELENAVGRADTGFVMTBEBELKLAVAARKGKRVVNTGVFDILHAGHVSYL 366  
DB 307 ELENAVGRADTGFVMTBEBELKLAVAARKGKRVVNTGVFDILHAGHVSYL 366  
QY 367 GDRILIVAVNSDASTRLKGSRPVNPLEFORMITVGALEAVDVMVVSFEEDTPORL 426  
DB 367 GDRILIVAVNSDASTRLKGSRPVNPLEFORMITVGALEAVDVMVVSFEEDTPORL 426





OY 184 ERYGATATLTNTLSFEFAVVGCKTEEBEIVEEGMKIAQ-YELISALLVYRSEOGMSLLQP 242  
 Db 180 SKYSHASLITTRNRALEQALHKKDLSHNLVKALQTLQETHTYHAMPVLVTLSEGGIAFLK 239  
 OY 243 GKAPLHMTQAOEYVDVTGAGPTVYGVLAATLTAAGNSLEBACFPANAAAGVYKGTST 302  
 Db 240 GEL-VNCTPIAKKVVDTGAGDTVIALSLTLSSLESKSLBACEFANAAAAVVGKMSAL 298  
 OY 303 VSPLELENAVGRADTGFVGMTEBELKTLVAAPAKRKEKVMNMGVDILHAGVSYLAN 362  
 Db 299 AS---LSEIALITLNGTHRKILPEKL---LETLENNQKQILVFNNGCFDILKHHASLYLQK 352  
 OY 363 ARKLGDRLLIVAVNSDASTRLKGDSPVNPLEQRMIVLGALEAVDWMVVSFEEDTPQRLIA 422  
 Db 353 AKALGDIILVVGINSNSIKRLKGDKRPIVSEKDRAPLASFCDVDVYVVGEDTPPIKLQ 412  
 OY 423 GILPDLVLKGGDYKPEBIAGSKRWANGGEVVLVLPNEDGCGSTNTIIKKIQ 473  
 Db 413 ALKPEDILKAGADYLNKEVYIGSE---LAKETRLIRBEIGTSALILEKIKR 459

RESULT 11  
 G84936  
 kinase [imported] - Buchnera sp. (strain APS)  
 C:Species: Buchnera sp.  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #ext\_change 09-May-2004  
 C:Accession: G84936  
 R:Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.  
 A:Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp.  
 A:Reference number: AB4930; MUID:20445173; PMID:10993077  
 A:Accession: G84936  
 A:Status: Preliminary

A:	Cross-references:	UNIPARC:UPI0000058442; GB:AP000398; GSPDB:GN00144
A:	Experimental source:	strain AFS
C:	Genetics:	
A:	Gene:	yB3052; BU060
C:	Superfamily:	r1Doxkinase
Query Match	33.9%; Score 811; DB 2; length 315;	
Best Local Similarity	50.0%; Pred. No. 8,7e-44;	
Matches 156; Conservative	63; Mismatches 93; Indels 0; Gaps 0;	
Gy	1 MKVTLPEEREGAVMYGVDMIDRYWVGPTSRISPEAPVPVYKNTIEEERGGAANVMANI	60
Db	1 MKKKLINFNNSSVLVAAGDILDECTYSKNHYMSEDLTPVLPNKIKEDGGAAANVAKNI	60
Gy	61 ASLGANAALVGLTGVIDAARALKSLADVNVCDFVSVPHTPIITTKRLVLSRNQQLRLD	120
Db	61 ABIGGYTIVFTGMDNEGLILKKLMDHIRIDSLLSIKKNNKITTYKRILRISEKKOLLRVD	120
Gy	121 FEEEGEGVDPQLHHRINQALLSIGALVYSDVAKGLASVQQIQTOLARKGAVPLTIDPKG	180
Db	121 FQEKYSKNNMLHQKIIDSISPKILVSDYAKGLIANQNIIDLAKKNSIFPLIDPKG	180

[illegible]

[illegible]

Db 194 GATLITPBARLKEVGVSKNENDLTEKAQNLRRHLDLTALILTRSEGMTLPSEGR-PI 252  
QY 248 HMPFOAOEVDYTGAGTIVGLAATLAAGNSLEACFPANAAAGVVGKLGTSVSPIE 307  
Db 253 YQPTFAOEVDVDSGAGDTVIAKGLIAGCTMPEAMYLANTAGVVAALGTFVCSFAR 312  
QY 308 LENA VRGRA 316  
Db 313 LTKALSGQS 321

## RESULT 15

T34841

probable bifunctional synthase /transferase - Streptomyces coelicolor

C/Species: Streptomyces coelicolor

C/Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 09-Jul-2004

C/Accession: T34841

R/Oliver, K.; Harrie, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, February 1999

A/Reference number: Z21559

A/Accession: T34841

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Release: 1-463 &lt;OL1&gt;

A/Cross-references: UNIPROT:Q9Z5B5, UNIPARC:UPI00000DAEC2, EMBL:AL035478; PIDN:CAB36595.

A/Experimental source: strain A3(2)

C/Genetic8:

C/Supfamily: hypothetical protein b3052

Query Match 24.8%; Score 594; DB 2; Length 463;

Best Local Similarity 36.3%; Pred. No. 6.5e-30;

Matches 173; Conservative 71; Mismatches 189; Indels 44; Gaps 13;

QY 9 ERAGVWVGVMDLDRVYGPTRISPEAPVVPVVTIERPGGAANVAMNIALGANAR 68  
Db 10 DRPLVVGADALDRDLTGTRADRLAPDAPVVGECABRIRPGGALAAVLAARDGREVT 69  
QY 69 LVGLTGIDDAARALSKLADVNNKCDVSVPTHTI-TKLAVLSRNOQLRLDPESGPEG 127  
Db 70 LIAGVGEDPAGLALRELIAP---WLKLIPLTGTVEKTRVLAQDRPVVRLDRGGG--- 123  
QY 128 VDPQPLHERINQALSSIG---ALVLSYAKGALASVQMIQLAKAGVPLIDPKGTFE 184  
Db 124 ----RVBAETDEARDALGCARAVLVSDYGRGADALRDVL---AARPLVMDPHRGP 175  
QY 185 RYRGATLLTNLSEFEAVNGKCKTE-----BEIVERGKMLADVELSALLVTRSEGM 237  
Db 176 PVGSTRLVTP--AEKA-HGFAPSEGRPGGLRAAALNAAALVMDRVAAVTVTLGSRG- 231  
QY 238 SLLOPGKAPLHMPFOAOEVDYTGAGTIVGLAATLAAGNSLEACFPANAAAGVVGK 297  
Db 232 ALLSYGHPPLVPAAPAHGDSGAGDRFAPATAAGLADGALVGEAVEGAVGATAFVAA 291  
QY 298 LGSTVSPLEBNVVRGRADTPFGVMTBBLKLAVAAARKGE--KYVMTNVGFIDLHAG 355  
Db 292 GGAANVPASGERALALPTD-----DQALAAIRAEKGVAAAGCCFDLHAG 342  
QY 356 HVSYLANARKLGDRLIVVNSDASTKRLKDSRPVNPLEQMTVLGALBAVDWVVSPEED 415  
Db 343 HVGILQAARLGDCLVVCNSDASVRKGGKGRPVNPLADRVYLRALACVDAVAVPDED 402  
QY 416 TPQRLINGILPDLVKGDIYKEBI--AGSKVWANGEVLVNFPDGCSTNIKK 470  
Db 403 TPERLLGELRPDVVWKGDYAGADLPBAGLLKEW--GQAVLPLPYLDGRSSTALAR 457

Search completed: March 18, 2006, 06:07:13  
Job time : 41 secs

**THIS PAGE LEFT BLANK**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: March 18, 2006, 05:59:28 ; Search time 231 Seconds  
(without alignments)  
1456.870 Million cell updates/sec

Title: US-10-520-820-13  
Predict score: 2393  
Sequence: 1 MKYTLDFRAGGVGVGDVW.....FEDGCTTNLIKXIQDQDKG 477

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues  
Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Uniprot 05.80.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2393	100.0	477	1	HLDE_ECOLI
2	2385	99.7	477	1	HLDE_ECOLI
3	2385	99.7	477	1	HLDE_ECOLI
4	2385	99.7	477	1	HLDE_ECOLI
5	2265	94.7	477	1	HLDE_SALTY
6	2265	94.7	477	1	HLDE_SALTY
7	2265	94.7	477	1	HLDE_SALTY
8	2264	94.6	477	2	OSPC86_SALPA
9	2094	87.5	476	2	OSPC86_SALPA
10	2081	87.0	476	2	OSPC86_SALPA
11	2076	86.8	476	2	OSPC86_SALPA
12	1949	81.4	474	1	HLDE_YERPE
13	1728	72.2	475	1	HLDE_PHOHL
14	1727	72.2	476	2	OSPC86_SALPA
15	1712	71.5	476	2	OSPC86_SALPA
16	1711	71.5	476	2	OSPC86_SALPA
17	1701	71.1	476	2	OSPC86_SALPA
18	1682	70.3	476	2	OSPC86_SALPA
19	1671	69.8	476	1	HLDE_VIBVU
20	1666	69.6	476	1	HLDE_VIBVU
21	1656	69.2	476	1	HLDE_VIBVU
22	1644	68.7	475	1	HLDE_VIBVU
23	1614	67.1	476	2	OSPC86_SALPA
24	1461.5	61.1	476	1	HLDE_SHOHL
25	1392	58.2	479	1	HLDE_CAMP
26	1385.5	57.9	474	2	OSPC86_SALPA
27	1361.5	56.9	474	2	OSPC86_SALPA
28	1350.5	56.7	474	1	HLDE_PSEBM
29	1350.5	56.4	474	1	HLDE_PSEBM
30	1350.5	56.4	474	1	HLDE_PSEBM
31	1337.5	55.9	477	2	OSPC86_SALPA

RESULT 1	HLDE_ECOLI	STANDARD	PRT	477 AA
32	1309.5	54.7	473	2
33	1305	54.5	475	1
34	955	39.9	470	1
35	934	39.0	326	2
36	918	38.4	479	2
37	904.5	37.8	483	1
38	891	37.2	474	1
39	884.5	37.0	490	1
40	882.5	36.9	494	1
41	868.5	36.3	457	2
42	865	36.1	482	1
43	863	36.1	461	1
44	863	36.1	496	1
45	855	35.7	461	2
ALIGNMENTS				
HLDE_ECOLI	STANDARD	PRT	477 AA	
AC	P7658			
DT	15-JUN-1999 (Rel. 38, Last sequence update)			
DT	15-JUN-1999 (Rel. 38, Last sequence update)			
DT	10-MAY-2005 (Rel. 47, Last annotation update)			
DE	Bifunctional protein hldR (Includes D-beta-D-heptose 7-phosphate kinase (EC 2.7.1.1) (D-beta-D-heptose 7-phosphotransferase); D-beta-D-heptose 1-phosphate adenylyltransferase (EC 2.7.7.-))			
GN	Name=hldR; Synonyms=rfab; waag; OrderedLocustNames=b3052;			
OS	Bacteriella coli.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Bacteriella.			
OX	NCBI_TaxID=562;			
OX	NCBI_TaxID=562;			
RP	NCBIOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].			
RC	STRAIN=K12 / MG1655;			
RX	MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;			
RA	Blaetner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,			
RA	Riley M., Colado-Vides J., Glaeser J.D., Koe C.K., Mayhew G.F.,			
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,			
RA	Mau B., Shao Y.;			
RT	"The complete genome sequence of Bacteriella coli K-12.";			
RL	Science 277:1453-1474(1997).			
RL	[2]			
RP	DOMAINS.			
RC	STRAIN=K12 / W1485;			
RX	MEDLINE=20096694; PubMed=10629197; DOI=10.1128/JB.182.2.488-497.2000;			
RA	Valvano M.A., Marolda C.L., Bittner M., Glaslin-Clay M., Simon T.L.,			
RA	Klena J.D.;			
RT	"The rfa gene from Bacteriella coli encodes a bifunctional protein			
RT	involved in the biosynthesis of the lipopolysaccharide core precursor			
RT	ADP-L-glycero-D-manno-heptose.";			
RL	J. Bacteriol. 182:488-497(2000).			
RL	[3]			
RP	FUNCTION. AND ADP-L-BETA-D-HEPTOSE BIOSYNTHESIS PATHWAY.			
RC	STRAIN=K12 / MG1655;			
RX	PubMed=11751812; DOI=10.1128/JB.184.2.363-369.2002;			
RA	Kreidinger B., Marolda C., Granger M., Zamyatina A., McArthur F.,			
RA	Kosma P., Valvano M.A., Messner P.;			
RT	"Biosynthesis pathway of ADP-L-glycero-beta-D-manno-heptose in			
RT	Bacteriella coli.";			
RL	J. Bacteriol. 184:363-369(2002).			
RL	J. Bacteriol. 184:363-369(2002).			
CC	-1- FUNCTION: Catalyzes the phosphorylation of D-glycero-D-manno-			
CC	heptose 7-phosphate at the C-1 position to form D,D-heptose-1,7-			
CC	bisphosphate.			
CC	-1- FUNCTION: Catalyzes the ADP transfer to D-glycero-D-manno-heptose			
CC	-1- CATALYTIC ACTIVITY: ATP + D-glycero-beta-D-manno-heptose 7-			
CC	phosphate = ADP + D-glycero-beta-D-manno-heptose 1,7-bisphosphate.			
CC	-1- CATALYTIC ACTIVITY: ATP + D-glycero-beta-D-manno-heptose 1-			
CC	phosphate = diphosphate + ADP-D-glycero-beta-D-manno-heptose.			
CC	-1- PATHWAY: Inner core lipopolysaccharide biosynthesis; ADP-L-			

```

CC glycerol-beta-D-manno-heptose biosynthesis; second step.
CC -1- PTHWAY: inner core lipopolysaccharide biosynthesis; ADP-L-
CC glycerol-beta-D-manno-heptose biosynthesis; fourth step.
CC -1- MISCELLANEOUS: In both reactions the enzyme functions only with
CC beta anomers.
CC -1- MISCELLANEOUS: The function of the domain II is independent from
CC the activity mediated by domain I.
CC -1- SIMILARITY: in the N-terminal section; belongs to the carbohydrate
CC kinase pfkB family.
CC -1- SIMILARITY: in the C-terminal section; belongs to the
CC cytidylyltransferase family.
-----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
-----
CC EMBL; U00096; AAC76088.1; -; Genomic_DNA.
CC PIR; B65093; B65093.
CC HSSP; P27623; 1COZ.
CC Echobase; EB3192; -.
CC Ecogene; EG13416; hide.
CC HAMAP; MF_01603; -; 1.
CC InterPro; IPR004820; Cyt trans rel.
CC InterPro; IPR004820; Cytidyltransferase.
CC InterPro; IPR002173; pfkB.
CC InterPro; IPR011611; pfkB_region.
CC InterPro; IPR011913; RfaE_dom_I.
CC InterPro; IPR011914; RfaE_dom_II.
CC Pfam; PF01467; CTP_transf_2; 1.
CC Pfam; PF00294; pfkB; 1.
CC TIGRFAMs; TIGR00125; cyt tran rel; 1.
CC TIGRFAMs; TIGR02198; rfaE_dom_I; 1.
CC TIGRFAMs; TIGR02199; rfaE_dom_II; 1.
CC PROSITE; PS00583; PFKB_KINASES_1; 1.
CC PROSITE; PS00584; PFKB_KINASES_2; FALSE_NEG.
CC Carbohydrate metabolism; Complete proteome; Kinase;
CC Lipopolysaccharide biosynthesis; Multifunctional enzyme;
CC Nucleotidyltransferase; Transferase.
CC REGION 1 318 Ribokinase.
CC REGION 344 477 Cytidylyltransferase.
CC SEQUENCE 477 AA; 51051 MM; 0F03CBE160B95389 CRC64;
SQ
Query Match 100.0%; Score 2393; DB 1; Length 477;
Best Local Similarity 100.0%; Pred. No. 2,2e-134;
Matches 477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Db 361 ANARKLGDRLIVAVNSDASTKRLKGDSPVNPVLEORMIVLGALEAVDWWVSPFEDTPQRL 420
Qy 421 IAGILPDLVKGDDYKPEEPIAGSKYEWANGBYLVNPFEDGCTTIKKIQQDDKXG 477
Db 421 IAGILPDLVKGDDYKPEEPIAGSKYEWANGBYLVNPFEDGCTTIKKIQQDDKXG 477

RESULT 2
HIDE ECOL6
ID HIDE_ECOL6 STANDARD; PRT; 477 AA.
AC Q8FDH5;
DT 25-OCT-2004 (Rel. 45, Created)
DT 25-OCT-2004 (Rel. 45, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Bifunctional protein hldE [includes: D-beta-D-heptose 7-phosphate
DE kinase (EC 2.7.1.1-) (D-beta-D-heptose 7-phosphotransferase); D-beta-D-
DE heptose 1-phosphate adenylyltransferase (EC 2.7.7.-)].
GN Name=hldE; Synonyms=rfaE; OrderedLocustNames=c3800;
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=O6:HL / CFT073 / ATCC 700928 / UPEC.
RX MEDLINE=22388234; PubMed=12471157; DOI=10.1073/pnas.252529799;
RA Welch R.A., Burdland V., Plunkett G., Iir, Redford P., Roesch P.,
RA Rayko D., Buckles E.L., Lion S.-R., Boutin A., Hackett J., Stroud D.,
RA Mewhew G.F., Rose D.U., Zhou S., Schwartz D.C., Ferna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli."
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
CC -1- FUNCTION: Catalyzes the phosphorylation of D-glycero-D-manno-
CC heptose 7-phosphate at the C-1 position to form D,D-heptose-1,7-.
CC bisphosphate (By similarity).
CC -1- FUNCTION: Catalyzes the ADP transfer to D-glycero-D-manno-heptose
CC 1-phosphate, yielding ADP-D,D-heptose (By similarity).
CC -1- CATALYTIC ACTIVITY: ATP + D-glycero-beta-D-manno-heptose 7-
CC phosphate = ADP + D-glycero-beta-D-manno-heptose 1,7-bisphosphate.
CC -1- CATALYTIC ACTIVITY: ATP + D-glycero-beta-D-manno-heptose 1-
CC phosphate = diphosphate + ADP-D-glycero-beta-D-manno-heptose.
CC -1- PTHWAY: Inner core lipopolysaccharide biosynthesis; ADP-L-
CC glycerol-beta-D-manno-heptose biosynthesis; second step.
CC -1- PTHWAY: Inner core lipopolysaccharide biosynthesis; ADP-L-
CC glycerol-beta-D-manno-heptose biosynthesis; fourth step.
CC -1- SIMILARITY: in the N-terminal section; belongs to the carbohydrate
CC kinase pfkB family.
CC -1- SIMILARITY: in the C-terminal section; belongs to the
CC cytidylyltransferase family.
-----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
-----
CC EMBL; AB016767; AN82245.1; -; Genomic_DNA.
CC HSSP; P27623; 1COZ.
CC HAMAP; MF_01603; -; 1.
CC InterPro; IPR004820; Cyt trans rel.
CC InterPro; IPR004820; Cytidylyltransferase.
CC InterPro; IPR002173; pfkB.
CC InterPro; IPR011611; pfkB_region.
CC InterPro; IPR011913; RfaE_dom_I.
CC InterPro; IPR011914; RfaE_dom_II.
CC Pfam; PF01467; CTP_transf_2; 1.
CC Pfam; PF00294; pfkB; 1.
CC TIGRFAMs; TIGR00125; cyt tran rel; 1.
CC TIGRFAMs; TIGR02198; rfaE_dom_I; 1.
CC TIGRFAMs; TIGR02199; rfaE_dom_II; 1.
CC PROSITE; PS00583; PFKB_KINASES_1; 1.

```

DR PROSITE; PS00584; PKB KINASES 2; FALSE NEG.  
 KW Carbohydrate metabolism; Complete proteome; Kinase;  
 KW Lipopolysaccharide biosynthesis; Multifunctional enzyme;  
 KW Nucleoside diphosphate kinase.  
 FT REGION 1 318 Ribokinase.  
 FT REGION 344 477 Cytidylyltransferase.  
 SO SEQUENCE 477 AA; 5106 MW; 0F03DBE171495389 CRC64;

Query Match 99.7%; Score 2386; DB 1; Length 477;  
 Best Local Similarity 99.8%; Pred. No. 5.8e-134;  
 Matches 476; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKTLPFEFBAQVWVGDVWLDKRYWGPTSRISPEAPVPPVKNVTEERPGGANVAMNI 60  
 DB 1 MKTLPFEFBAQVWVGDVWLDKRYWGPTSRISPEAPVPPVKNVTEERPGGANVAMNI 60

QY 61 ASIGANARLVGLTGIDDAARALSKSLADVNVKCDPVSPHPITTKLRVLSRNOQLRLD 120  
 DB 61 ASIGANARLVGLTGIDDAARALSKSLADVNVKCDPVSPHPITTKLRVLSRNOQLRLD 120

QY 121 FEEFEGVDPPQPLHERINQALSSIGALVLSDYAKGALASVQOMIQARKAGVPLIDPKG 180  
 DB 121 FEEFEGVDPPQPLHERINQALSSIGALVLSDYAKGALASVQOMIQARKAGVPLIDPKG 180

QY 121 FEEFEGVDPPQPLHERINQALSSIGALVLSDYAKGALASVQOMIQARKAGVPLIDPKG 180  
 DB 121 FEEFEGVDPPQPLHERINQALSSIGALVLSDYAKGALASVQOMIQARKAGVPLIDPKG 180

QY 181 TDFERYGATLLTNNLSPEAVGCKTEERIVRGMKLTADVELSALLVTRSEQMSLL 240  
 DB 181 TDFERYGATLLTNNLSPEAVGCKTEERIVRGMKLTADVELSALLVTRSEQMSLL 240

QY 241 QPGKAPLHPTQAOEVVDVTGAGDTVIGVLAATLAAGNSLEACFPANAAAGVVGKLG 300  
 DB 241 QPGKAPLHPTQAOEVVDVTGAGDTVIGVLAATLAAGNSLEACFPANAAAGVVGKLG 300

QY 301 STVSPLEENAVRGRADTGFVMTSEEBIKALVAARRGKGVNTNGVFDILAHGHSYL 360  
 DB 301 STVSPLEENAVRGRADTGFVMTSEEBIKALVAARRGKGVNTNGVFDILAHGHSYL 360

QY 361 ANAKKLDRLIVANNSDASTRLKDGSPVNPLOHMTIVGALRAVWVWVFEEDTPORL 420  
 DB 361 ANAKKLDRLIVANNSDASTRLKDGSPVNPLOHMTIVGALRAVWVWVFEEDTPORL 420

QY 421 IAGILPDLVKGSDYKKEBEIAGSKVWANGEVVLNFEDECSYTNIIKKIQDDK 477  
 DB 421 IAGILPDLVKGSDYKKEBEIAGSKVWANGEVVLNFEDECSYTNIIKKIQDDK 477

RESULT 3  
 HIDE EC057 STANDARD, PRT, 477 AA.  
 AC 07A07; O8XBW4;  
 DT 25-OCT-2004 (Rel. 45, Last sequence update)  
 DT 25-OCT-2004 (Rel. 48, Last annotation update)  
 DE 13-SEP-2005 (Rel. 48, Last annotation update)  
 DE Multifunctional protein hldc (includes: D-beta-D-heptose 7-phosphate  
 kinase (EC 2.7.1.-) (D-beta-D-heptose 7-phosphotransferase); D-beta-D-  
 heptose 1-phosphate adenosyltransferase (EC 2.7.7.-)).  
 GN Name=hldc; Synonym=erfAB; Ordered locusName=24405; ECs3935;  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 NCBI\_Taxid=83334;  
 RX NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927 / EHEC;  
 RX MEDLINE=21074935; PubMed=1120551; DOI=10.1038/35054089;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glaesner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Postel G., Heckett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RL "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";  
 RL Nature 409:529-533 (2001).  
 RN [2]

RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=O157:H7 / Sakai / RIMD 0509952 / EHEC;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayaishi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
 RA Kohara S., Shiba T., Hattori M., Shingawa H.;  
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
 O157:H7 and genomic comparison with a laboratory strain K-12.";  
 RT DNA Res. 8:11-22 (2001).  
 RL "FUNCTION: Catalyzes the phosphorylation of D-glycero-D-manno-  
 heptose 7-phosphate at the C-1 position to form D,D-heptose-1,7-  
 bisphosphate (By similarity).  
 CC -1- FUNCTION: Catalyzes the ADP transfer to D-glycero-D-manno-heptose  
 1-phosphate, yielding ADP-D,D-heptose (By similarity).  
 CC -1- CATALYTIC ACTIVITY: ATP + D-glycero-beta-D-manno-heptose 7-  
 phosphate = ADP + D-glycero-beta-D-manno-heptose 1,7-bisphosphate.  
 CC -1- CATALYTIC ACTIVITY: ATP + D-glycero-beta-D-manno-heptose 1-  
 phosphate = diphosphate + ADP-D-glycero-beta-D-manno-heptose.  
 CC -1- PATHWAY: Inner core lipopolysaccharide biosynthesis; ADP-L-  
 glycero-beta-D-manno-heptose biosynthesis; second step.  
 CC -1- PATHWAY: Inner core lipopolysaccharide biosynthesis; ADP-L-  
 glycero-beta-D-manno-heptose biosynthesis; fourth step.  
 CC -1- SIMILARITY: In the N-terminal section; belongs to the carbohydrate  
 kinase pfkB family.  
 CC -1- SIMILARITY: In the C-terminal section; belongs to the  
 cytidylyltransferase family.  
 CC -----  
 CC This Swiss-Prot entry is copyrighted. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use as long as its content is in no way modified and this statement is not  
 removed.  
 CC -----  
 CC EMBL; AB005174; AAC58186.1; -; Genomic DNA.  
 CC EMBL; BA000007; BAB37358.1; -; Genomic DNA.  
 DR PIR; F85965; F85965.  
 DR PIR; G91120; G91120.  
 DR HSSP; P27623; 1COZ.  
 DR HAMAP; MF\_01603; -; 1.  
 DR InterPro; IPR004821; Cyt. trans. rel.  
 DR InterPro; IPR004820; Cytidylyltransf.  
 DR InterPro; IPR002173; pfkB.  
 DR InterPro; IPR011611; pfkB region.  
 DR InterPro; IPR011913; RfaB\_dom\_1.  
 DR InterPro; IPR011914; RfaB\_dom\_11.  
 DR Pfam; PF01467; CTP\_transf\_2; 1.  
 DR Pfam; PF00294; pfkB; 1.  
 DR TIGRFAMs; TIGR00125; Cyt. tran. rel.; 1.  
 DR TIGRFAMs; TIGR02198; rfaB\_dom\_11; 1.  
 DR TIGRFAMs; TIGR02199; rfaB\_dom\_11; 1.  
 DR PROSITE; PS00583; PKB KINASES 1; 1.  
 DR PROSITE; PS00584; PKB KINASES 2; FALSE NEG.  
 KW Carbohydrate metabolism; Complete proteome; Kinase;  
 KW Lipopolysaccharide biosynthesis; Multifunctional enzyme;  
 KW Nucleoside diphosphate kinase.  
 FT REGION 1 318 Ribokinase.  
 FT REGION 344 477 Cytidylyltransferase.  
 SO SEQUENCE 477 AA; 5106 MW; BB877FEF656B67C CRC64;

Query Match 99.7%; Score 2385; DB 1; Length 477;  
 Best Local Similarity 99.6%; Pred. No. 6.6e-134;  
 Matches 475; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKTLPFEFBAQVWVGDVWLDKRYWGPTSRISPEAPVPPVKNVTEERPGGANVAMNI 60  
 DB 1 MKTLPFEFBAQVWVGDVWLDKRYWGPTSRISPEAPVPPVKNVTEERPGGANVAMNI 60

QY 61 ASIGANARLVGLTGIDDAARALSKSLADVNVKCDPVSPHPITTKLRVLSRNOQLRLD 120  
 DB 61 ASIGANARLVGLTGIDDAARALSKSLADVNVKCDPVSPHPITTKLRVLSRNOQLRLD 120

QY 121 FEEFEGVDPPQPLHERINQALSSIGALVLSDYAKGALASVQOMIQARKAGVPLIDPKG 180  
 DB 121 FEEFEGVDPPQPLHERINQALSSIGALVLSDYAKGALASVQOMIQARKAGVPLIDPKG 180

```

Db 121 FEESFEGVDPOPLHERINQALSSIGALVLSDYAKGALASVQOMIQLARKAGVPLIDPKG 180
Qy 181 TDFERRYGATLLTPNLSFEFAVVKCKTEBEIYERGMKLADYELSLALVTRSEQMSLL 240
Db 181 TDFERRYGATLLTPNLSFEFAVVKCKTEBEIYERGMKLADYELSLALVTRSEQMSLL 240
Qy 241 QPKAPLPHMTQAOEYVDVYDAGDTVIGVLAATLAAGNSLEBACFPANAAAGVVGKLG 300
Db 241 QPKAPLPHMTQAOEYVDVYDAGDTVIGVLAATLAAGNSLEBACFPANAAAGVVGKLG 300
Qy 301 STVSPIELENAVRGADTGFVMTBEELKLAVAARKGGEVVMVNGVFDILAHGHSYL 360
Db 301 STVSPIELENAVRGADTGFVMTBEELKLAVAARKGGEVVMVNGVFDILAHGHSYL 360
Qy 361 ANARKLGDRLIVAVNSDASTKRLKGDSPVNPLEORMIVLGALSAVDVWVSFEEDTQRL 420
Db 361 ANARKLGDRLIVAVNSDASTKRLKGDSPVNPLEORMIVLGALSAVDVWVSFEEDTQRL 420
Qy 421 IAGILPDLVKGDDYKPEEIASKEVWANGEVVLNFEDEGCTTNIKKIQDDKKG 477
Db 421 IAGILPDLVKGDDYKPEEIASKEVWANGEVVLNFEDEGCTTNIKKIQDDKKG 477

```

```

RESULT 4
HIDE SHIFL
ID HIDE SHIFL STANDARD; PRT; 477 AA.
AC 07UBI8; 083047;
DT 25-OCT-2004 (Rel. 45, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Bifunctional protein hldE (includes: D-beta-D-heptose 7-phosphate
DE kinase (EC 2.7.1.-) (D-beta-D-heptose 7-phosphotransferase); D-beta-D-
DE heptose 1-phosphate adenosyltransferase (EC 2.7.7.-)).
GN Name:hldE; Synonyms:rfaB; Ordered locus names=SF3093, S3289;
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP NUCLEOTIDE SEQUENCE (LARGE SCALE GENOMIC DNA).
RC STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12384590; DOI=10.1093/nar/gkf566;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Chang H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157."
RT Nucleic Acids Res. 30:4432-4441 (2002).
RL [2]
RN NUCLEOTIDE SEQUENCE (LARGE SCALE GENOMIC DNA).
RC STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=22590274; PubMed=12704152;
RA DOI=10.1128/JAI.71.5.2775-2786.2003;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.W., Deng W.,
RA Fourtner G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
RT flexneri serotype 2a strain 2457T."
RL Infect. Immun. 71:2775-2786 (2003).
CC -1- FUNCTION: Catalyzes the phosphorylation of D-glycero-D-manno-
CC heptose 7-phosphate at the C-1 position to form D,D-heptose-1,7-
CC biphosphate (By similarity).
CC -1- FUNCTION: Catalyzes the ADP transfer to D-glycero-D-manno-heptose
CC 1-phosphate, yielding ADP-D,D-heptose (By similarity).
CC -1- CATALYTIC ACTIVITY: ATP + D-glycero-beta-D-manno-heptose 7-
CC phosphate = ADP + D-glycero-beta-D-manno-heptose 1,7-bisphosphate.
CC -1- CATALYTIC ACTIVITY: ATP + D-glycero-beta-D-manno-heptose 1-
CC phosphate = diphosphate + ADP-D-glycero-beta-D-manno-heptose.
CC -1- PATHWAY: Inner core lipopolysaccharide biosynthesis; ADP-L-

```

```

CC glycero-beta-D-manno-heptose biosynthesis; second step.
CC -1- PATHWAY: Inner core lipopolysaccharide biosynthesis; ADP-L-
CC glycero-beta-D-manno-heptose biosynthesis; fourth step.
CC -1- SIMILARITY: In the N-terminal section, belongs to the carbohydrate
CC kinase pfk family.
CC -1- SIMILARITY: In the C-terminal section, belongs to the
CC cytidylyltransferase family.
CC -----
CC This Swiss-Prot entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL; AB005674; AAN44569.1; -; Genomic DNA.
CC EMBL; AB016988; AAP18381.1; -; Genomic DNA.
CC HSSP; P27623; 1COZ.
CC DR HAMAP; MF_01603; -; 1.
CC DR InterPro; IPR004821; Cyt_trans_rel.
CC DR InterPro; IPR004820; Cytidylyltransf.
CC DR InterPro; IPR002173; PfkB.
CC DR InterPro; IPR011611; PfkB_region.
CC DR InterPro; IPR011913; RfaE_dom_1.
CC DR InterPro; IPR011914; RfaE_dom_11.
CC DR Pfam; PF01467; CTP_transfer_2; 1.
CC DR Pfam; PF00294; PfkB; 1.
CC DR TIGRPFAMs; TIGR00125; Cyt tran rel; 1.
CC DR TIGRPFAMs; TIGR02198; rfaE_dom_1; 1.
CC DR TIGRPFAMs; TIGR02199; rfaE_dom_11; 1.
CC DR PROSITE; PS00583; PFKB_KINASES_1; 1.
CC DR PROSITE; PS00584; PFKB_KINASES_2; FALSE_NEG.
CC KW Carbohydrate metabolism; Complete proteome; Kinase;
CC KW Lipopolysaccharide biosynthesis; Multifunctional enzyme;
CC KW Nucleotidyltransferase; Transferase.
CC FT REGION 1 318 Rbokinase.
CC FT REGION 344 477 Cytidylyltransferase.
CC FT CONFLICT 133 133 M -> L (in Ref. 2).
CC SQ SEQUENCE 477 AA; 51129 MW; E0F224053402A28 CRC64;

Query Match 99.7%; Score 2385; DB 1; Length 477;
Best Local Similarity 99.6%; Pred. No. 6 6e-134; Indels 0; Gaps 0;
Matches 475; Conservative 1; Mismatches 1;

Qy 1 MKYTLPEFERAGVWVGVNLDRTYWGPTSRISPEAPVPVYKXNTIEERPGAAVAMNI 60
Db 1 MKYTLPEFERAGVWVGVNLDRTYWGPTSRISPEAPVPVYKXNTIEERPGAAVAMNI 60
Qy 61 ASIGANARLVGLTGIDDAARALSKSLADVWVKCDPVSPHPTITTKLRVTSRQQLRLD 120
Db 61 ASIGANARLVGLTGIDDAARALSKSLADVWVKCDPVSPHPTITTKLRVTSRQQLRLD 120
Qy 121 FEESFEGVDPOPLHERINQALSSIGALVLSDYAKGALASVQOMIQLARKAGVPLIDPKG 180
Db 121 FEESFEGVDPOPLHERINQALSSIGALVLSDYAKGALASVQOMIQLARKAGVPLIDPKG 180
Qy 181 TDFERRYGATLLTPNLSFEFAVVKCKTEBEIYERGMKLADYELSLALVTRSEQMSLL 240
Db 181 TDFERRYGATLLTPNLSFEFAVVKCKTEBEIYERGMKLADYELSLALVTRSEQMSLL 240
Qy 241 QPKAPLPHMTQAOEYVDVYDAGDTVIGVLAATLAAGNSLEBACFPANAAAGVVGKLG 300
Db 241 QPKAPLPHMTQAOEYVDVYDAGDTVIGVLAATLAAGNSLEBACFPANAAAGVVGKLG 300
Qy 301 STVSPIELENAVRGADTGFVMTBEELKLAVAARKGGEVVMVNGVFDILAHGHSYL 360
Db 301 STVSPIELENAVRGADTGFVMTBEELKLAVAARKGGEVVMVNGVFDILAHGHSYL 360
Qy 361 ANARKLGDRLIVAVNSDASTKRLKGDSPVNPLEORMIVLGALSAVDVWVSFEEDTQRL 420
Db 361 ANARKLGDRLIVAVNSDASTKRLKGDSPVNPLEORMIVLGALSAVDVWVSFEEDTQRL 420
Qy 421 IAGILPDLVKGDDYKPEEIASKEVWANGEVVLNFEDEGCTTNIKKIQDDKKG 477
Db 421 IAGILPDLVKGDDYKPEEIASKEVWANGEVVLNFEDEGCTTNIKKIQDDKKG 477

```



DB 421 IAGILPDLVKGDPYKBEINGSKEVWANGGEVLVNFEDGCTTNIKKIQODKKG 477

RESULT 5

HIDE\_SALTY STANDARD; PRT; 477 AA.

AC OXHEM9; Q7AM91; (Rel. 45, Last sequence update)

DT 25-OCT-2004 (Rel. 45, Last sequence update)

DT 10-MAY-2005 (Rel. 47, Last sequence update)

DE Bifunctional protein hldc [includes: D-beta-D-heptose 7-phosphate kinase (EC 2.7.1.-) (D-beta-D-heptose 7-phosphotransferase); D-beta-D-heptose 1-phosphate adenosyltransferase (EC 2.7.7.-)]

GN Name=ldc; Synonym=rfab; OrderedLocName=STW3339; t3320;

OS Salmonella typhimurium.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.

OC NCBI\_TaxID=601;

RA NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].

RC STRAIN=CT18;

RX MEDLINE=21534947; PubMed=11677608; DOI=10.1038/35101607;

RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J., Churcher C.M., Mungall K.L., Bentley S.D., Holden M.T.G., Sebatina M., Baker S., Bauman D., Brooks K., Chillingworth T., Connerton P., Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J., Kretzschmar T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K., Krogan A., Larsen T.S., Leather S., Moule S., O'Garra P., Parry C., Quail M.A., Rutherford K.M., Simmonds M., Skelton J., Stevens K., Whitehead S., Barrrell B.G.;

RT "Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium CT18."

RL Nature 413:848-852(2001).

RA NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].

RC STRAIN=Ty2 / ATCC 700931;

RX MEDLINE=22531367; PubMed=12644504;

RX DOI=10.1128/JB.185.7.2330-2337.2003;

RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J., Burland V., Kodymani V., Schwartz D.C., Blattner F.R.;

RT "Comparative genomics of Salmonella enterica serovar Typhimurium Ty2 and CT18."

RL J. Bacteriol. 185:2330-2337(2003).

CC -1- FUNCTION: Catalyzes the phosphorylation of D-glycero-D-manno-heptose 7-phosphate at the C-1 position to form D,D-heptose-1,7-bisphosphate (By similarity).

CC -1- FUNCTION: Catalyzes the ADP transfer to D-glycero-D-manno-heptose 1-phosphate, yielding ADP-D,D-heptose (By similarity).

CC -1- CATALYTIC ACTIVITY: ATP + D-glycero-beta-D-manno-heptose 7-phosphate = ADP + D-glycero-beta-D-manno-heptose 1,7-bisphosphate.

CC -1- CATALYTIC ACTIVITY: ATP + D-glycero-beta-D-manno-heptose 1-phosphate = diphosphate + ADP-D-glycero-beta-D-manno-heptose.

CC -1- PATHWAY: Inner core lipopolysaccharide biosynthesis; ADP-L-glycero-beta-D-manno-heptose biosynthesis; second step.

CC -1- PATHWAY: Inner core lipopolysaccharide biosynthesis; ADP-L-glycero-beta-D-manno-heptose biosynthesis; fourth step.

CC -1- SIMILARITY: In the N-terminal section; belongs to the carbohydrate kinase pfkB family.

CC -1- SIMILARITY: In the C-terminal section; belongs to the cytidyltransferase family.

CC This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.

CC EMBL; AL637278; CAD07725.1; -; Genomic DNA.

CC EMBL; AB016844; AA070663.1; -; Genomic DNA.

CC HSBP; P27623; IC02.

CC HAMAP; MF\_01603; -; 1.

CC InterPro; IPR004821; Cyt trans-rel.

CC InterPro; IPR004820; Cytidyltransferase.

DR InterPro; IPR002173; pfkB.

DR InterPro; IPR011611; pfkB region.

DR InterPro; IPR011913; RfaE\_dom\_1.

DR InterPro; IPR011914; RfaE\_dom\_11.

DR Pfam; PF01467; CTP\_transf\_2; 1.

DR Pfam; PF00294; pfkB; 1.

DR TIGRFAMs; TIGR00125; cyt tran rel; 1.

DR TIGRFAMs; TIGR02198; rfaE\_dom\_1; 1.

DR TIGRFAMs; TIGR02199; rfaE\_dom\_11; 1.

DR PROSITE; PS00583; PFKB\_KINASES\_1; 1.

DR PROSITE; PS00584; PFKB\_KINASES\_2; FALSE NEG.

KW Carbohydrate metabolism; Complete proteome; Kinase;

KW Lipopolysaccharide biosynthesis; Multifunctional enzyme;

KW Nucleotidyltransferase; Transferase.

FT REGION 1 318 Ribokinase.

FT REGION 344 477 Cytidyltransferase.

SQ SEQUENCE 477 AA; 51124 MW; E4F6B1DEB0568C CRC64;

Query Match 94.7%; Score 2265; DB 1; Length 477;

Best Local Similarity 93.5%; Pred. No. 9.3e-127;

Matches 445; Conservative 19; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKYTLPEFBAQVWVGVNLDYRWYGPTRISPEAPVPVKNVTIEERPGAAVAMNT 60

DB 1 MKVNLPAFERAGVWVGVNLDYRWYGPTRISPEAPVPVKNVTIEERPGAAVAMNI 60

QY 61 ASIGANARLVGLTGIDDAARALSKLADVNVKCDPVSPVPHPTTKLRVSRNOQLRLD 120

DB 61 ASIGANARLVGLTGIDDAARALSKLADVNVKCDPVSPVPHPTTKLRVSRNOQLRLD 120

QY 121 PEEGEGVDPPQPHERRINQALSSIGALVSDYAKGALASVQOMIQARAKAVPLIDPKG 180

DB 121 PEEGEGVDPPQPHERRINQALSSIGALVSDYAKGALASVQOMISLARQGVPLIDPKG 180

QY 181 TDFERRYGATLLTPNISEFAVAVGCKTEBIEVERGKLIADYELSLAVTRSEOGNSLL 240

DB 181 TDFERRYGATLLTPNISEFAVAVGCKSEBELVERGKLIADYELSLAVTRSEOGNTLL 240

QY 241 QPGKAPLHMTQAOEYVDVTGADTVYGVLAATLAAGNSLEKCFPANAAGVVGGLCT 300

DB 241 QPKAPLHMTQAOEYVDVTGADTVYGVLAATLAAGNTLEBAKCFPANAAGVVGGLCT 300

QY 301 STSPFIEBNAVGRADPTGFWTTEBELKLAVAARRGKGVMTNGVFPDILAHGVSYL 360

DB 301 STSPFIEBNAVGRADPTGFWTTEBELKLAVAARRGKGVMTNGVFPDILAHGVSYL 360

QY 361 ANARKIGDRLLIVANSDSATKRLKGSRPVNPLEQRMIVLGALBAVDVWVSFEEDTPQL 420

DB 361 ANARKIGDRLLIVANSDSATKRLKGSRPVNPLEQRMIVLGALBAVDVWVSFEEDTPQL 420

QY 421 IAGILPDLVKGDPYKBEINGSKEVWANGGEVLVNFEDGCTTNIKKIQODK 476

DB 421 IAGILPDLVKGDPYKBEINGSKEVWANGGEVLVNFEDGCTTNIKKIQODE 476

RESULT 6

HIDE\_SALTY STANDARD; PRT; 477 AA.

AC Q7CPR9; Q9AJ74; Q9RFY7; Q9RFY8;

DT 25-OCT-2004 (Rel. 45, Created)

DT 25-OCT-2004 (Rel. 45, Last sequence update)

DT 10-MAY-2005 (Rel. 47, Last sequence update)

DE Bifunctional protein hldc [includes: D-beta-D-heptose 7-phosphate kinase (EC 2.7.1.-) (D-beta-D-heptose 7-phosphotransferase); D-beta-D-heptose 1-phosphate adenosyltransferase (EC 2.7.7.-)]

GN Name=ldc; Synonym=rfab; waab; OrderedLocName=STW3320;

OS Salmonella typhimurium.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.

OC NCBI\_TaxID=602;

RA NUCLEOTIDE SEQUENCE, AND ROLE IN LPS BIOSYNTHESIS.

RC PubMed=12441667; DOI=10.1023/A:1021103501626;



DR EMBL; AB017220; AA67053.1; -; Genomic\_DNA.

KM Complete proteome; Transferrase.

Seq SEQUENCE 477 AA; 51124 MM; BAF6B1DBE80566C CRC64;  
Query Match 94.7%; Score 2265; DB 2; Length 477;  
Best Local Similarity 93.5%; Pred. No. 9,36-127;  
Matches 445; Conservative 19; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKVTLPEFEPAGVNVGDVMDLDRWYGTPTSRISPEAPVPPVKNVTTEERPGGANVAMNI 60  
DB 1 MKVTLPEFEPAGVNVGDVMDLDRWYGTPTSRISPEAPVPPVKNVTTEERPGGANVAMNI 60  
QY 61 ASLGANARLVGLTGDIDDAARLSKSLADVNVKCPFSVPHPTITTKLRVSRNOQLRLD 120  
DB 61 ASLGANARLVGLTGDIDDAARLSKSLADVNVKCPFSVPHPTITTKLRVSRNOQLRLD 120  
QY 121 FEEGFEVDPPQPHERRINQALSGISGALVLSDYAKGALTSVQTMISLARQGVFLIDPKG 180  
DB 121 FEEGFEVDPPQPHERRINQALSGISGALVLSDYAKGALTSVQTMISLARQGVFLIDPKG 180  
QY 181 TDPERYRGATLTLPNLSPEFAVVGKCTEERIVRGKCLADYELSLALVTRSEQMSLL 240  
DB 181 TDPERYRGATLTLPNLSPEFAVVGKCTEERIVRGKCLADYELSLALVTRSEQMSLL 240  
QY 241 QPKAPLHPTQAOEYVDVYAGDVTYGVLAATLAAGNSLEBACFPANAAAGVVGDLGT 300  
DB 241 QPKAPLHPTQAOEYVDVYAGDVTYGVLAATLAAGNSLEBACFPANAAAGVVGDLGT 300  
QY 301 STVSPIELBNVNRGRADTGFVMTTEBELKLAVAARRGKGVMTNGVFDILHAGVSYL 360  
DB 301 STVSPIELBNVNRGRADTGFVMTTEBELKLAVAARRGKGVMTNGVFDILHAGVSYL 360  
QY 361 ANARKGDRLLIVANSDSASTKRLKGSRPVNPLEQRMTVIGALEAVDVMVSFEEDTPQRL 420  
DB 361 ANARKGDRLLIVANSDSASTKRLKGSRPVNPLEQRMTVIGALEAVDVMVSFEEDTPQRL 420  
QY 421 IAGILPDLVKGDDYKPEBIAGSKVWANGGEVVLNFEDECSITNIIKKIQODKK 476  
DB 421 IAGILPDLVKGDDYKPEBIAGSKVWANGGEVVLNFEDECSITNIIKKIQODKK 476

# RESULT 8

Q5PC86\_SALPA PRELIMINARY; PRT; 477 AA.

AC Q5PC86;  
DT 01-FEB-2005 (TrEMBLrel. 29, Created)  
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)  
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)  
DE ADP-heptose synthase.  
GN Name: rfaF, Ordered locus names=SPA3068;  
OS Salmonella paratyphi-a.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmonella.  
OX NCBI\_TaxID=54388;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=ATCC 9150;  
RX PubMed=15531882; DOI=10.1038/ng1470;  
RA Mclelland M., Sanders R.E., Clifton S.W., Latreille P.,  
RA Porwollik S., Sabo A., Meyer R., Bieri T., Ozesky P., Mclelland M.,  
RA Harkins C.R., Wang C., Nguyen C., Berghoff A., Elliott G.,  
RA Kohlberg S., Strong C., Du F., Carter J., Krenitzki C., Layman D.,  
RA Leonard S., Sun H., Fulton L., Nash W., Miner T., Mink P., Florea L.,  
RA Deleahunty K., Fronick C., Magrini V., Nhan M., Warren W., Florea L.,  
RA Spleth J., Wilson R.K.;  
RT "Comparison of genome degradation in Paratyphi A and Typhi, human-  
RT restricted serovars of Salmonella enterica that cause typhoid.";  
RL Nat. Genet. 36:1268-1274(2004).  
DR EMBL; CP000026; AA078903.1; -; Genomic\_DNA.  
DR GO; GO:0016301; P:kinase activity; IEA.  
DR GO; GO:0016779; F:nucleotidyltransferase activity; IEA.  
DR GO; GO:0016740; F:transferase activity; IEA.  
DR GO; GO:0009058; P:biogenesis; IEA.

DR GO; GO:0005975; P:carbohydrate metabolism; IEA.

DR GO; GO:0009103; P:lipopolysaccharide biosynthesis; IEA.

DR InterPro; IPR004820; Cytidylyltransf.

DR InterPro; IPR004821; Cyt. trans. rel.

DR InterPro; IPR002173; PfkB.

DR InterPro; IPR011611; PfkB region.

DR InterPro; IPR011913; RfaE dom. I.

DR InterPro; IPR011914; RfaE dom. II.

DR Pfam; PF01467; CTP\_transf\_2; 1.

DR Pfam; PF00294; PfkB; 1.

DR TIGRPFAMs; TIGR00125; Cyt. tran. rel; 1.

DR TIGRPFAMs; TIGR02198; rfaE dom. I; 1.

DR TIGRPFAMs; TIGR02199; rfaE dom. II; 1.

DR PROSITE; PS00583; PFKB\_KINASES\_1; 1.

KM Complete proteome.

Seq SEQUENCE 477 AA; 51138 MM; 3F5C6B1DFC02C69E CRC64;

Query Match 94.6%; Score 2264; DB 2; Length 477;  
Best Local Similarity 93.3%; Pred. No. 1,1e-126;  
Matches 444; Conservative 20; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKVTLPEFEPAGVNVGDVMDLDRWYGTPTSRISPEAPVPPVKNVTTEERPGGANVAMNI 60  
DB 1 MKVTLPEFEPAGVNVGDVMDLDRWYGTPTSRISPEAPVPPVKNVTTEERPGGANVAMNI 60  
QY 61 ASLGANARLVGLTGDIDDAARLSKSLADVNVKCPFSVPHPTITTKLRVSRNOQLRLD 120  
DB 61 ASLGANARLVGLTGDIDDAARLSKSLADVNVKCPFSVPHPTITTKLRVSRNOQLRLD 120  
QY 121 FEEGFEVDPPQPHERRINQALSGISGALVLSDYAKGALTSVQTMISLARQGVFLIDPKG 180  
DB 121 FEEGFEVDPPQPHERRINQALSGISGALVLSDYAKGALTSVQTMISLARQGVFLIDPKG 180  
QY 181 TDPERYRGATLTLPNLSPEFAVVGKCTEERIVRGKCLADYELSLALVTRSEQMSLL 240  
DB 181 TDPERYRGATLTLPNLSPEFAVVGKCTEERIVRGKCLADYELSLALVTRSEQMSLL 240  
QY 241 QPKAPLHPTQAOEYVDVYAGDVTYGVLAATLAAGNSLEBACFPANAAAGVVGDLGT 300  
DB 241 QPKAPLHPTQAOEYVDVYAGDVTYGVLAATLAAGNSLEBACFPANAAAGVVGDLGT 300  
QY 301 STVSPIELBNVNRGRADTGFVMTTEBELKLAVAARRGKGVMTNGVFDILHAGVSYL 360  
DB 301 STVSPIELBNVNRGRADTGFVMTTEBELKLAVAARRGKGVMTNGVFDILHAGVSYL 360  
QY 361 ANARKGDRLLIVANSDSASTKRLKGSRPVNPLEQRMTVIGALEAVDVMVSFEEDTPQRL 420  
DB 361 ANARKGDRLLIVANSDSASTKRLKGSRPVNPLEQRMTVIGALEAVDVMVSFEEDTPQRL 420  
QY 421 IAGILPDLVKGDDYKPEBIAGSKVWANGGEVVLNFEDECSITNIIKKIQODKK 476  
DB 421 IAGILPDLVKGDDYKPEBIAGSKVWANGGEVVLNFEDECSITNIIKKIQODKK 476

# RESULT 9

O6D164\_ERWCT PRELIMINARY; PRT; 476 AA.

AC O6D164;  
DT 25-OCT-2004 (TrEMBLrel. 28, Created)  
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)  
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
DE ADP-heptose synthase (EC 2.7.-.-).  
GN Name: rfaF, Synonym=vaab; Ordered locus names=BCA3584;  
OS Erwinia carotovora (subsp. atroseptica) (Pectobacterium atrosepticum).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Pectobacterium.  
OX NCBI\_TaxID=29471;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RC STRAIN=SCRI 1043 / ATCC BAA-672;  
RX PubMed=15263089; DOI=10.1073/pnas.0402424101;  
RA Bell K.S., Seibalt M., Pritchard L., Holden M.T.G., Hyman L.J.,  
RA Holve M.C., Thomson N.R., Bentley S.D., Churcher L.J.C., Mungall K.,

RA Atkin R., Bason N., Brooks K., Chillingworth T., Clark K., Doggett J.,  
 RA Fraser A., Hance Z., Hauser H., Jagels K., Moule S., Norbertczak H.,  
 RA Ormond D., Price C., Quail M.A., Sanders M., Walker D., Whitehead S.,  
 RA Salmond G.P.C., Birch P.R.J., Parkhill J., Toth I.K.,  
 RT "Genome sequence of the enterobacterial phytopathogen *Erwinia*  
 RT *carotovora* subsp. *atroseptica* and characterization of virulence  
 RT factors.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 101:11105-11110(2004).  
 DR EMBL: BX950851; CAG76482.1; -; Genomic DNA.  
 DR GO: GO:0016301; P:kinase activity; IEA.  
 DR GO: GO:0016779; F:nucleotidyltransferase activity; IEA.  
 DR GO: GO:0016740; F:transferase activity; IEA.  
 DR GO: GO:0009058; P:biosynthesis; IEA.  
 DR GO: GO:0009103; P:carbohydrate metabolism; IEA.  
 DR GO: GO:0009103; P:lipopolysaccharide biosynthesis; IEA.  
 DR InterPro: IPR004820; Cytidylyltransf.  
 DR InterPro: IPR004821; Cyt\_transf.  
 DR InterPro: IPR002173; PfkB.  
 DR InterPro: IPR011611; PfkB\_region.  
 DR InterPro: IPR011913; RfaE\_dom\_1.  
 DR InterPro: IPR011914; RfaE\_dom\_1.  
 DR Pfam: PF01467; CTP\_transf\_2; 1.  
 DR Pfam: PF00294; PfkB; 1.  
 DR TIGRFAMs: TIGR00125; Cyt\_tran\_rel; 1.  
 DR TIGRFAMs: TIGR02198; rfaE\_dom\_1; 1.  
 DR TIGRFAMs: TIGR02199; rfaE\_dom\_1; 1.  
 DR PROSITE: PS00583; PFKB\_KINASES\_1; 1.  
 KM Complete proteome; Transferase.  
 SQ SEQUENCE 476 AA; 50763 MW; 19A32257B56B0DB CRC64;

Query Match 87.5%; Score 2094; DB 2; Length 476;  
 Best Local Similarity 87.0%; Pred. No. 1,4e-116;  
 Matches 409; Conservative 30; Mismatches 31; Indels 0; Gaps 0;  
 QY 3 VTLEPEFBAAGVWVGDVMDLDRYWGPTSRISPEAPVPVVKVNTTIEERGGAAVAMNIAS 62  
 DB 1 MTLDPFRAGVAVVGDVMDLDRYWGPTSRISPEAPVPVVKVNTTIEERGGAAVAMNIAS 60  
 QY 63 LGANARLVGLTGIDDAARALSKSLADVNVKCDPVSVPHTPTTKLRVLSRNOQLRLDPE 122  
 DB 61 LGASRLVGLTGVDAAARALAKGENVVKCDPVSVPHTPTTKLRVLSRNOQLRLDPE 120  
 QY 123 EGFEFVDPOPLHERINQALSSIGALVLSDVAKGALASVQOMIQLARKAGVPLIDPKGT 182  
 DB 121 EGFEFVDPOPLHERINQALSSIGALVLSDVAKGALASVQOMIQLARKAGVPLIDPKGT 180  
 QY 183 FERRRGATLTPNISEFAVVGKCTEERIEYERGMKLIADYELSLALVTRSEQMSLLQ 242  
 DB 181 FSRRGATLTPNISEFAVVGKCTEERIEYERGMKLIADYELSLALVTRSEQMSLLQ 240  
 QY 243 GKAPLHMPTOAEVVDVAGDVTYGVLAATLAAGNSLEBACFPANAAAGVVGKLGST 302  
 DB 241 GKAPLHMPTOAEVVDVAGDVTYGVLAATLAAGNSLEBACFPANAAAGVVGKLGST 300  
 QY 303 VSPLEENAVGRADTGVTGTEEBELKLAVAARKGKGVMTNGVPDILHAGVSYLAN 362  
 DB 301 VSPLEENAVGRADTGVTGTEEBELKLAVAARKGKGVMTNGVPDILHAGVSYLAN 360  
 QY 363 ARKAGDRLIVAVNSDASTKRLKGSRPVNPLEQRMIVGALBAVDWVVSFEEDTPQRL 422  
 DB 361 ARKAGDRLIVAVNSDASTKRLKGSRPVNPLEQRMIVGALBAVDWVVSFEEDTPQRL 420  
 QY 423 GILPDLVKGGDYKPEEISKEVWANGGEVLVNFEGGCTNNIKKIQ 472  
 DB 421 GILPDLVKGGDYKPEEISKEVWANGGEVLVNFEGGCTNNIKKIQ 470

RESULT 10  
 Q65V3 YERPS PRELIMINARY; PRT; 476 AA.  
 AC Q65V3\_2004 (TREMBlrel. 28, Created)  
 DT 25-OCT-2004 (TREMBlrel. 28, Last sequence update)

DT 25-OCT-2004 (TREMBlrel. 28, Last annotation update)  
 DE ADP-heptose synthase (EC 2.7.-.-).  
 GN Name=rfab; Synonyms=waab; OrderedLocustNames=YPTB3407;  
 OS Yersinia pseudotuberculosis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Yersinia.  
 OX NCBI\_TaxID=633;  
 RN [1]  
 RP NCITBOTIDE SEQUENCE (LARGE SCALE GENOMIC DNA).  
 RC STRAIN=IP32953 / Serotype I;  
 RX PubMed=15358858; DOI=10.1073/pnas.040401201;  
 RA Challa P.S.G., Carmel E., Larimer F.W., Lemerding J., Stoutland P.O.,  
 RA Ragala W.M., Georgescu A.M., Vergez L.M., Land M.L., Motin V.L.,  
 RA Brubaker R.R., Fowler J., Hinnelbusch J., Marceau M., Médigue C.,  
 RA Simonet M., Chantal-Francoise V., Souza B., Dacheux D., Elliott J.M.,  
 RA Darbise A., Hauser J.J., Garcia E.;  
 RT "Insights into the evolution of *Yersinia* pests through whole-genome  
 RT comparison with *Yersinia pseudotuberculosis*.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 101:13826-13831(2004).  
 DR EMBL: BX936398; CAH22645.1; -; Genomic DNA.  
 DR GO: GO:0016779; F:nucleotidyltransferase activity; IEA.  
 DR GO: GO:0016740; F:transferase activity; IEA.  
 DR GO: GO:0009058; P:biosynthesis; IEA.  
 DR InterPro: IPR004820; Cytidylyltransf.  
 DR InterPro: IPR004821; Cyt\_trans\_rel.  
 DR InterPro: IPR002173; PfkB.  
 DR InterPro: IPR011611; PfkB\_region.  
 DR InterPro: IPR011913; RfaE\_dom\_1.  
 DR InterPro: IPR011914; RfaE\_dom\_1.  
 DR Pfam: PF01467; CTP\_transf\_2; 1.  
 DR Pfam: PF00294; PfkB; 1.  
 DR TIGRFAMs: TIGR00125; Cyt\_tran\_rel; 1.  
 DR TIGRFAMs: TIGR02198; rfaE\_dom\_1; 1.  
 DR TIGRFAMs: TIGR02199; rfaE\_dom\_1; 1.  
 DR PROSITE: PS00583; PFKB\_KINASES\_1; 1.  
 KM Complete proteome; Transferase.  
 SQ SEQUENCE 476 AA; 51202 MW; B56550CA2D31726D CRC64;

Query Match 87.0%; Score 2081; DB 2; Length 476;  
 Best Local Similarity 86.3%; Pred. No. 8,5e-116;  
 Matches 410; Conservative 32; Mismatches 33; Indels 0; Gaps 0;  
 QY 1 MKVTLPEFBAAGVWVGDVMDLDRYWGPTSRISPEAPVPVVKVNTTIEERGGAAVAMNI 60  
 DB 1 MKVTLPEFBAAGVAVVGDVMDLDRYWGPTSRISPEAPVPVVKVNTTIEERGGAAVAMNI 60  
 QY 61 ASIGANARLVGLTGIDDAARALSKSLADVNVKCDPVSVPHTPTTKLRVLSRNOQLRLD 120  
 DB 61 ASIGANARLVGLTGIDDAARALSKSLADVNVKCDPVSVPHTPTTKLRVLSRNOQLRLD 120  
 QY 121 FERRRGATLTPNISEFAVVGKCTEERIEYERGMKLIADYELSLALVTRSEQMSLLQ 240  
 DB 121 FERRRGATLTPNISEFAVVGKCTEERIEYERGMKLIADYELSLALVTRSEQMSLLQ 240  
 QY 243 GKAPLHMPTOAEVVDVAGDVTYGVLAATLAAGNSLEBACFPANAAAGVVGKLGST 300  
 DB 241 GKAPLHMPTOAEVVDVAGDVTYGVLAATLAAGNSLEBACFPANAAAGVVGKLGST 300  
 QY 303 VSPLEENAVGRADTGVTGTEEBELKLAVAARKGKGVMTNGVPDILHAGVSYLAN 360  
 DB 301 VSPLEENAVGRADTGVTGTEEBELKLAVAARKGKGVMTNGVPDILHAGVSYLAN 360  
 QY 363 ARKAGDRLIVAVNSDASTKRLKGSRPVNPLEQRMIVGALBAVDWVVSFEEDTPQRL 420  
 DB 361 ARKAGDRLIVAVNSDASTKRLKGSRPVNPLEQRMIVGALBAVDWVVSFEEDTPQRL 420  
 QY 423 GILPDLVKGGDYKPEEISKEVWANGGEVLVNFEGGCTNNIKKIQ 475  
 DB 421 GILPDLVKGGDYKPEEISKEVWANGGEVLVNFEGGCTNNIKKIQ 475

RESULT 11  
HLDE\_YERPE STANDARD; PRT; 476 AA.  
AC Q82160; Q74RQ3; O8C2S0;  
DT 25-OCT-2004 (Rel. 45, Created)  
DT 25-OCT-2004 (Rel. 45, Last sequence update)  
DT 10-MAY-2005 (Rel. 47, Last annotation update)  
DE Bifunctional protein hlde [includes: D-beta-D-heptose 7-phosphate  
kinase (EC 2.7.1.-) (D-beta-D-heptose 7-phosphotransferase); D-beta-D-  
heptose 1-phosphate adenylyltransferase (EC 2.7.7.-)].  
GN Name=hlde; Synonyms=rfab, waab;  
OS OrderedLocustNames=YPO0554, y3524, YP2965;  
OC Yersinia pestis.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Yersinia.  
OX NCBI\_TaxId=632;  
[1]  
RA NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RP STRAIN=CO-92 / Biovar Orientalis;  
RC MEDLINE=21470413; PubMed=11586360; DOI=10.1038/35097083;  
RA Parkhill J., Wren B.W., Thomson N.R., Tlball R.W., Holden M.T.G.,  
RA Prentice M.B., Sebatia M., James K.D., Church C.M., Mungall K.L.,  
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.-M.,  
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,  
RA Leathley S., Moule S., Oyston P.C.F., Quail M.A., Rutherford K.M.,  
RA Almonde M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;  
RT "Genome sequence of Yersinia pestis, the causative agent of plague."  
RL Nature 433:523-527(2001).  
[2]  
RA NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RP STRAIN=KIM5 / Biovar Mediaevalis;  
RC MEDLINE=22137863; PubMed=12142430;  
RA DOI=10.1126/SCIENCE.1184.16.4601-4611.2002;  
RA Deng W., Burdall V., Plunkett G. III, Boutin A., Mayhew G.F., Lise P.,  
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,  
RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,  
RA Straley S.C., McDonough K.A., Nilles M.L., Watson J.S., Blattner F.R.,  
RA Perry R.D.;  
RT "Genome sequence of Yersinia pestis KIM."  
RL J. Bacteriol. 184:4601-4611(2002).  
[3]  
RA NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RP STRAIN=91001 / Biovar Mediaevalis;  
RC PubMed=15368893;  
RA Song Y., Tong Z., Wang J., Wang L., Guo Z., Han Y., Zhang J., Pei D.,  
RA Zhou D., Qin H., Pang X., Han Y., Zhai J., Li M., Cui B., Qi Z.,  
RA Jin L., Dai R., Chen F., Li S., Ye C., Du Z., Lin W., Wang J., Yu J.,  
RA Yang H., Wang J., Huang P., Yang R.;  
RT "Complete genome sequence of Yersinia pestis strain 91001, an isolate  
avirulent to humans."  
RL DNA Res. 11:179-197(2004).  
-1- FUNCTION: Catalyzes the phosphorylation of D-glycero-D-manno-  
heptose 7-phosphate at the C-1 position to form D,D-heptose-1,7-  
bisphosphate (By similarity).  
CC -1- FUNCTION: Catalyzes the ADP transfer to D-glycero-D-manno-heptose  
1-phosphate, yielding ADP-D,D-heptose (By similarity).  
CC -1- CATALYTIC ACTIVITY: ATP + D-glycero-beta-D-manno-heptose 7-  
phosphate = ADP + D-glycero-beta-D-manno-heptose 1,7-bisphosphate.  
CC -1- CATALYTIC ACTIVITY: ATP + D-glycero-beta-D-manno-heptose 1-  
phosphate = diphosphate + ADP-D-glycero-beta-D-manno-heptose.  
CC -1- PATHWAY: Inner core lipopolysaccharide biosynthesis; ADP-L-  
glycero-beta-D-manno-heptose biosynthesis; second step.  
CC -1- PATHWAY: Inner core lipopolysaccharide biosynthesis; ADP-L-  
glycero-beta-D-manno-heptose biosynthesis; fourth step.  
CC -1- SIMILARITY: In the N-terminal section, belongs to the  
kinase pfkb family.  
CC -1- SIMILARITY: In the C-terminal section, belongs to the  
cyclidlyltransferase family.  
CC This Swiss-Prot entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use as long as its content is in no way modified and this statement is not  
CC removed.  
CC  
CC EMBL; AJ141444; CAC89508.1; -; Genomic DNA.  
CC EMBL; AB013955; AAM87072.1; ALT INIT; Genomic DNA.  
CC EMBL; AB017138; AAS63148.1; ALT\_INIT; Genomic DNA.  
CC F01; A10080; A10080.  
CC HSP; P27623; 1C02.  
CC HAMAP; MF 01603; -; 1.  
CC InterPro: IPR004821; Cyt trans rel.  
CC InterPro: IPR004820; Cytidyltransferase.  
CC InterPro: IPR002173; Pfkb.  
CC InterPro: IPR011611; Pfkb region.  
CC InterPro: IPR011913; Rfab\_dom\_1.  
CC InterPro: IPR011914; Rfab\_dom\_11.  
CC Pfam; PF01467; CTP transef\_2; 1.  
CC Pfam; PF00294; Pfkb; 1.  
CC TIGRfam; TIGR00125; Cyt tran rel; 1.  
CC TIGRfam; TIGR02198; Rfab\_dom\_1; 1.  
CC TIGRfam; TIGR02199; Rfab\_dom\_11; 1.  
CC PROSITE; PS00583; PFKB\_KINASES\_1; 1.  
CC PROSITE; PS00584; PFKB\_KINASES\_2; FALSE NEG.  
CC Carbohydrate metabolism; Complete proteome; Kinase;  
CC Lipopolysaccharide biosynthesis; Multifunctional enzyme;  
CC Nucleotidyltransferase; Transferase.  
CC REGION 1 318 Ribokinae.  
CC FT REGION 344 476 Cytidyltransferase.  
CC ST SEQUENCE 476 AA; 5128 MW; E68FF382DC892636 CRC64;  
Query Match 86.8%; Score 2076; DB 1; Length 476;  
Best Local Similarity 86.1%; Pred. No. 1.7e-115;  
Matches 409; Conservative 32; Mismatches 34; Indels 0; Gaps 0;  
QY 1 MKTTLPEFERAGVGVVGVGDWVLDKRYWYPTSRISPEAPVPVVKVNTIEERPGANVAMNI 60  
DB 1 MKTTLPEFERAGVGVVGVGDWVLDKRYWYPTSRISPEAPVPVVKVNTIEERPGANVAMNI 60  
QY 61 ASIGAVARLVGLTGIDDAARALICLSSEVVRCDPVSPVTHPTTKLRVLSRNOQLRLD 120  
DB 61 ASIGAVARLVGLTGIDDAARALICLSSEVVRCDPVSPVTHPTTKLRVLSRNOQLRLD 120  
QY 121 FEESFGVDPPPIFEHQLQLPQIGALVSDVAKGALNSVQPMIQLARKANVVLIDPKG 180  
DB 121 FEESFGVDPPPIFEHQLQLPQIGALVSDVAKGALNSVQPMIQLARKANVVLIDPKG 180  
QY 181 TDFERYGATLTPNISEFAVAVGCKTEIEIYERQKLIADYELSLALVTRSEOGMSL 240  
DB 181 TDFERYGATLTPNISEFAVAVGCKTEIEIYERQKLIADYELSLALVTRSEOGMSL 240  
QY 181 SDFERRGATLTPNISEFAVAVGCKTEIEIYERQKLIADYELSLALVTRSEOGMSL 240  
DB 181 SDFERRGATLTPNISEFAVAVGCKTEIEIYERQKLIADYELSLALVTRSEOGMSL 240  
QY 241 QPKAPLMPFTQAOBYVDVYAGDTVIYGLAATLAAGNSIEBAQFPANAAAGVVGKLT 300  
DB 241 QPKAPLMPFTQAOBYVDVYAGDTVIYGLAATLAAGNSIEBAQFPANAAAGVVGKLT 300  
QY 241 QLKPKPLHPTQAKVFDVYAGDTVIYGLAATLAAGNSIEBAQFPANAAAGVVGKLT 300  
DB 241 QLKPKPLHPTQAKVFDVYAGDTVIYGLAATLAAGNSIEBAQFPANAAAGVVGKLT 300  
QY 301 STSPSTELKNAVGRADTGCVNTEIEIKLAAVAAKRGKVMYTMGVPDILAHGVSYL 360  
DB 301 STSPSTELKNAVGRADTGCVNTEIEIKLAAVAAKRGKVMYTMGVPDILAHGVSYL 360  
QY 361 ANARKIGDRILIVANSNDASKRLKGDSPVNPPLQSRNIVGALAEVAVVWVSPEDTPORL 420  
DB 361 ANARKIGDRILIVANSNDASKRLKGDSPVNPPLQSRNIVGALAEVAVVWVSPEDTPORL 420  
QY 421 IAGILPDLVKGGDYKPEBIAGSEKAVANGGEVLVINFEDGSGTTNIIKKIQDK 475  
DB 421 IAGILPDLVKGGDYKPEBIAGSEKAVANGGEVLVINFEDGSGTTNIIKKIQDK 475  
RESULT 12  
HLDE\_PROHL STANDARD; PRT; 474 AA.  
AC Q7NOC3;  
DT 25-OCT-2004 (Rel. 45, Created)

DT 25-OCT-2004 (Rel. 45, last sequence update)  
 DT 10-MAY-2005 (Rel. 47, last annotation update)  
 DE Bifunctional protein hldR [includes: D-beta-D-heptose 7-phosphate  
 DE kinase (EC 2.7.1.-) (D-beta-D-heptose 7-phosphotransferase); D-beta-D-  
 DE heptose 1-phosphate adenosyltransferase (EC 2.7.7.-)].  
 GN Name=hldR; Synonyms=rfas; OrderedLocNames=plus3968;  
 OS Bacteroides uniformis (subsp. laundoni).  
 OC Bacteroidetes; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Photorhabdus.  
 NCBI\_TaxID=141679;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=TT01;  
 RX MEDLINE=22957627; PubMed=14528314; DOI=10.1038/nbt886;  
 RA Duchaud E., Rusniok C., Frangeul L., Buchrieser C., Givaudan A.,  
 RA Taouit S., Bocs S., Bouraux-Eude C., Chandler M., Charles J.-F.,  
 RA Dassa E., Deroo R., Derzelle S., Freysinet G., Gaudriault S.,  
 RA Meligne C., Lancelis A., Powell K., Signier P., Vincent R., Wingate V.,  
 RA Zouine M., Glaser P., Boemare N., Danchin A., Kunst F.,  
 RT "The genome sequence of the entomopathogenic bacterium Photorhabdus  
 RT luminescens";  
 RT Nat. Biotechnol. 21:1307-1313(2003).  
 CC -1- FUNCTION: Catalyzes the phosphorylation of D-glycero-D-manno-  
 CC heptose 7-phosphate at the C-1 position to form D,D-heptose-1,7-  
 CC bisphosphate (By similarity).  
 CC -1- FUNCTION: Catalyzes the ADP transfer to D-glycero-D-manno-heptose  
 CC 1-phosphate, yielding ADP-D,D-heptose (By similarity).  
 CC -1- CATALYTIC ACTIVITY: ATP + D-glycero-beta-D-manno-heptose 7-  
 CC phosphate = ADP + D-glycero-beta-D-manno-heptose 1,7-bisphosphate.  
 CC -1- CATALYTIC ACTIVITY: ATP + D-glycero-beta-D-manno-heptose 1-  
 CC phosphate = diphosphate + ADP-D-glycero-beta-D-manno-heptose.  
 CC -1- PATHWAY: Inner core lipopolysaccharide biosynthesis; ADP-L-  
 CC glycero-beta-D-manno-heptose biosynthesis; second step.  
 CC -1- PATHWAY: Inner core lipopolysaccharide biosynthesis; ADP-L-  
 CC glycero-beta-D-manno-heptose biosynthesis; fourth step.  
 CC -1- SIMILARITY: In the N-terminal section; belongs to the carbohydrate  
 CC kinase pfkB family.  
 CC -1- SIMILARITY: In the C-terminal section; belongs to the  
 CC cytidyltransferase family.  
 CC -----  
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use as long as its content is in no way modified and this statement is not  
 CC removed.  
 CC -----  
 CC EMBL, BX571872; CAE16340.1; -, Genomic DNA.  
 CC Photobacterium plus3968; -.  
 DR HAMAP, MF\_01603; -, 1.  
 DR InterPro, IPR004821, Cyt trans rel.  
 DR InterPro, IPR004820, Cytidyltransferase.  
 DR InterPro, IPR002173, PEKB.  
 DR InterPro, IPR011611, PEKB region.  
 DR InterPro, IPR011813, RfaE\_dom\_1.  
 DR InterPro, IPR011914, RfaE\_dom\_11.  
 DR Pfam, PF01467, CYP transf\_2; 1.  
 DR Pfam, PF00294, PEKB; 1.  
 DR TIGRFAMs, TIGR00125, Cyt trans rel; 1.  
 DR TIGRFAMs, TIGR02198, rfaE\_dom\_1; 1.  
 DR TIGRFAMs, TIGR02199, rfaE\_dom\_11; 1.  
 DR PROSITE, PS00583, PEKB KINASES\_1; 1.  
 DR PROSITE, PS00584, PEKB KINASES\_2; FALSE NEG.  
 KW Carbohydrate metabolism; Complete proteome; Kinase;  
 KW Lipopolysaccharide biosynthesis; Multifunctional enzyme;  
 KW Nucleotidyltransferase; Transferase.  
 FT REGION 1 318 Ribokinase.  
 FT REGION 344 474 Cytidyltransferase.  
 SQ SEQUENCE 474 AA; 5130 MW; DCAC4E30D108862D CRC64;  
 Query Match 81.4%; Score 1949; DB 1; Length 474;  
 Best Local Similarity 81.1%; Pred. No. 6.2e-108;  
 Matches 383; Conservative 37; Mismatches 52; Indels 0; Gaps 0;

QY 1 MKVTLPEFERAGVGVGVVMDLDRWYGPTRISRPAPVPVVKVNTIEERPGAAVAMNTI 60  
 DB 1 MKVTLPEFERAGVGVGVVMDLDRWYGPTRISRPAPVPVVKVNTIEERPGAAVAMNTI 60  
 QY 61 ASIGANARLVGLTGIDDAARALKSLADVNVKCDPVSVPFHPTITLGRVTSRQQLRLD 120  
 DB 61 AALGANSNHLGLGVIDDAARALKSEKRSVYVRCDFVSPVPHPTVTKLRVTSRQQLRLD 120  
 QY 121 FEEGFEVDPPQPHERRINQALSSIGALVLSDYAKGALAAVQMIQLARKGVPVLLDPKG 180  
 DB 121 FEEGFEVDPPQPHERRINQALSSIGALVLSDYAKGALAAVQMIQLARKGVPVLLDPKG 180  
 QY 181 TDFERRRGATLLTPNISFEPAVVGKCTEERIVERGKLIADYLSALLVTSRQGSLL 240  
 DB 181 NDFERRRGATLLTPNISFEPAVVGKCTEERIVERGKLIADYLSALLVTSRQGSLL 240  
 QY 241 QPKAPLPMTOQOEYVDVYAGADTVYIGVLAATLAAAGNSLEACFPANAAAGVVGGLGT 300  
 DB 241 SVDPPLHPTQOEYVDVYAGADTVYIGVLAATLAAAGNSLEACFPANAAAGVVGGLGT 300  
 QY 301 STVSPIELENAVGRADTGFVMTTEBELKLAVAARKGKRVMTNGVPDILAHGVSYL 360  
 DB 301 STVSPIELENAVGRADTGFVMTTEBELKLAVAARKGKRVMTNGVPDILAHGVSYL 360  
 QY 361 ANARKGLRLIVANSDASTKRLKGSRPVNPPLRQMTVYALBAVWVVSFEEDTPQRL 420  
 DB 361 ENARKGLRLIVANSDASTKRLKGSRPVNPPLRQMTVYALBAVWVVSFEEDTPQRL 420  
 QY 421 IADILPDLVKGDDYKPEERFAGSKYEWANGGEVLVNFEDGCTTNIKKIQ 472  
 DB 421 IADILPDLVKGDDYKPEERFAGSKYEWANGGEVLVNFEDGCTTNIKKIQ 472  
 ID HIDE ACTPL STANDARD; PRT; 475 AA.  
 AC QGGLD;  
 DT 25-OCT-2004 (Rel. 45, last sequence update)  
 DT 10-MAY-2005 (Rel. 47, last annotation update)  
 DE Bifunctional protein hldR [includes: D-beta-D-heptose 7-phosphate  
 DE kinase (EC 2.7.1.-) (D-beta-D-heptose 7-phosphotransferase); D-beta-D-  
 DE heptose 1-phosphate adenosyltransferase (EC 2.7.7.-)].  
 GN Name=hldR; Synonyms=rfas;  
 OS Actinobacillus pleuropneumoniae (Haemophilus pleuropneumoniae).  
 OC Bacteroidetes; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 OC Pasteurellaceae; Actinobacillus.  
 NCBI\_TaxID=715;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [GENOMIC DNA], AND ROLE IN LPS BIOSYNTHESIS.  
 RC STRAIN=S 4074 / Serotype 1;  
 RX MEDLINE=22683163; PubMed=12798993; DOI=10.1016/S0378-1097(03)00247-7;  
 RA Provost M., Harel J., Labrie J., Sirolis M., Jacques M.,  
 RT "Identification, cloning and characterization of rfbE of  
 RT Actinobacillus pleuropneumoniae serotype 1, a gene involved in  
 RT lipopolysaccharide inner-core biosynthesis";  
 RT FEMS Microbiol. Lett. 223:7-14(2003).  
 RL FEMS Microbiol. Lett. 223:7-14(2003).  
 CC -1- FUNCTION: Catalyzes the phosphorylation of D-glycero-D-manno-  
 CC heptose 7-phosphate at the C-1 position to form D,D-heptose-1,7-  
 CC bisphosphate (Probable).  
 CC -1- FUNCTION: Catalyzes the ADP transfer to D-glycero-D-manno-heptose  
 CC 1-phosphate, yielding ADP-D,D-heptose (Probable).  
 CC -1- CATALYTIC ACTIVITY: ATP + D-glycero-beta-D-manno-heptose 7-  
 CC phosphate = ADP + D-glycero-beta-D-manno-heptose 1,7-bisphosphate.  
 CC -1- CATALYTIC ACTIVITY: ATP + D-glycero-beta-D-manno-heptose 1-  
 CC phosphate = diphosphate + ADP-D-glycero-beta-D-manno-heptose.  
 CC -1- PATHWAY: Inner core lipopolysaccharide biosynthesis; ADP-L-  
 CC glycero-beta-D-manno-heptose biosynthesis; second step.  
 CC -1- PATHWAY: Inner core lipopolysaccharide biosynthesis; ADP-L-  
 CC glycero-beta-D-manno-heptose biosynthesis; fourth step.  
 CC -1- SIMILARITY: In the N-terminal section; belongs to the carbohydrate  
 CC kinase pfkB family.  
 CC -1- SIMILARITY: In the C-terminal section; belongs to the





DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 10-MAY-2005 (Rel. 47, Last annotation update)  
 DE Bifunctional protein hIdb [includes: D-beta-D-heptose 7-phosphate  
 DE kinase (EC 2.7.1.-) (D-beta-D-heptose 7-phosphotransferase); D-beta-D-  
 DE heptose 1-phosphate adenosyltransferase (EC 2.7.7.-)]  
 DE Name=HdE; Synonyms=rtab; waab; OrderedLocustNames=H11526;  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 OC Pasteurellaceae; Haemophilus.  
 OC NCBI\_TaxId=727;  
 RX NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RX STRAIN=rd / KW20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirschner E.F.,  
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McEwen K., Sutton G.G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghegan N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 RT Rd.";  
 RL Science 269:496-512(1995).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE OF 1-427, AND ROLE IN LOS BIOSYNTHESIS.  
 RC STRAIN=NTM1 2019;  
 RX MEDLINE=95172727; PubMed=7868252;  
 RA Lee N.-G., Sunshine M.G., Apicella M.A.;  
 RT "Molecular cloning and characterization of the nontypeable Haemophilus  
 RT influenzae 2019 rfaE gene required for lipopolysaccharide  
 RT biosynthesis.";  
 RL Infect. Immun. 63:818-824(1995).  
 CC -1- FUNCTION: Catalyzes the phosphorylation of D-glycero-D-manno-  
 CC heptose 7-phosphate at the C-1 position to form D,D-heptose-1,7-  
 CC bisphosphate (Probable).  
 CC -1- FUNCTION: Catalyzes the ADP transfer to D-glycero-D-manno-heptose  
 CC 1-phosphate, yielding ADP-D,D-heptose (Probable).  
 CC -1- CATALYTIC ACTIVITY: ATP + D-glycero-beta-D-manno-heptose 7-  
 CC phosphate = ADP + D-glycero-beta-D-manno-heptose 1,7-bisphosphate.  
 CC -1- CATALYTIC ACTIVITY: ATP + D-glycero-beta-D-manno-heptose 1-  
 CC phosphate = diphosphate + ADP-D-glycero-beta-D-manno-heptose.  
 CC -1- PATHWAY: Lipopoligosaccharide (LOS) core biosynthesis; ADP-L-  
 CC glycero-beta-D-manno-heptose biosynthesis; second step.  
 CC -1- PATHWAY: Lipopoligosaccharide (LOS) core biosynthesis; ADP-L-  
 CC glycero-beta-D-manno-heptose biosynthesis; fourth step.  
 CC -1- SIMILARITY: In the N-terminal section; belongs to the carbohydrate  
 CC kinase pfkB family.  
 CC -1- SIMILARITY: In the C-terminal section; belongs to the  
 CC cytidyltransferase family.  
 CC -1- CAUTION: Ref.2 sequence differs from that shown due to a  
 CC frameshift in position 331.  
 CC  
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use as long as its content is in no way modified and this statement is not  
 CC removed.  
 CC  
 CC EMBL: U32828; AAC23172.1; -; Genomic DNA.  
 DR EMBL: U17642; AAC43516.1; ALT\_FRAME; Genomic DNA.  
 DR PIR: C64127; C64127.  
 DR HSP: P27623; 1COZ.  
 DR TIGR: H11526; -.  
 DR HAMAP: MF\_01603; -; 1.  
 DR InterPro: IPR004821; Cyt trans rel.  
 DR InterPro: IPR004820; Cytidylyltransf.  
 DR InterPro: IPR002173; pfkB.  
 DR InterPro: IPR011611; pfkB\_region.  
 DR InterPro: IPR011913; RfaE\_dom\_1.  
 DR InterPro: IPR011914; RfaE\_dom\_1i.

DR Pfam; PF01467; CTP transf. 2; 1.  
 DR Pfam; PF00294; pfkB; 1.  
 DR TIGRPFAMs; TIGR00125; Cyt tran rel; 1.  
 DR TIGRPFAMs; TIGR02198; rfaE\_dom\_1; 1.  
 DR TIGRPFAMs; TIGR02199; rfaE\_dom\_1i; 1.  
 DR PROSITE; PS00583; pfkB\_KINASES\_1; FALSE\_NEG.  
 DR PROSITE; PS00584; pfkB\_KINASES\_2; FALSE\_NEG.  
 DR Carbohydrate metabolism; Complete proteome; Kinase;  
 KM Multifunctional enzyme; Nucleotidyltransferase; Transferase.  
 FT REGION 1 318  
 FT REGION 344 476  
 FT CONFLICT 36 36 Cytidylyltransferase.  
 FT CONFLICT 84 84 A -> R (in Ref. 2).  
 FT CONFLICT 84 84 L -> H (in Ref. 2).  
 FT CONFLICT 170 170 A -> G (in Ref. 2).  
 FT CONFLICT 214 214 E -> K (in Ref. 2).  
 FT CONFLICT 273 273 A -> T (in Ref. 2).  
 FT CONFLICT 278 278 R -> C (in Ref. 2).  
 SQ SEQUENCE 476 AA; 51946 MW; 4F241C08D2C6951E CRC64;  
 Query Match 71.5%; Score 1712; DB 1; Length 476;  
 Best local Similarity 71.0%; Pred. No. 8.2e-94;  
 Matches 331; Conservative 58; Mismatches 77; Indels 0; Gaps 0;  
 QY 7 EFERAGVGVGVVMDLDRVYGTPTSRISPEAPVPVKNVTIEEPGGAANYANNIASIGAN 66  
 DB 7 EFKQAKVLVGVDMVLDRLVWFGATNRISPEAPVPVQVEBERAGGAANYANNIASINVP 66  
 QY 67 ARVLGLTGIDDAARALSKSLADVNVKDFVSPTHTPTTLRLVSSNQQLRIIDFEERGE 126  
 DB 67 VQLMGILGQDETSSALSLLEKOKIDCNFVALTHPTTLRLVSSNQQLRIIDFEEDFN 126  
 QY 127 GVPQPLHERINQALSSIGALVLDYAKGALASVQOMIQARRAGVPVLIDPKGTDFERY 186  
 DB 127 NVQCKQLALLESANVNYGALLISDYGKGLKVQGMQIQRKANVPVLIDPKGTDFERY 186  
 QY 187 RGATLTLPNLSFEFAVVGKCTEEIEVERGMKLIADYELISALIVTSEQMSILQPEKAP 246  
 DB 187 RGATLTLPNLSFEFAVVGKCTEEIEIEKGLKISDIEILTALVTRSEKMTLLRPVQEP 246  
 QY 247 LHMPTQAOEYVDYTGADPTVIGVLAATTLAAGNSLEBACFPANNAAGVVGKLTSTVSP 306  
 DB 247 YHLPTVAKEVFDYTGADPTVISVLAALDGRFESSCYLANVAGLVGKLTSTVSTV 306  
 QY 307 ELENAVGRADTFGVTTEBELKLAAVAARKRGKVTMTNGVDDIILHAGVSYLANARKL 366  
 DB 307 ELENAIHARETFGIMSEBELKDAVAQAKAREKIVMTNGVDDIILHGHISYLENARKL 366  
 QY 367 GDRLLIVANSDASTKRLKGDSPVNPLEORMIVLGALEAVDVVVSFEEDTPQRLIAGILP 426  
 DB 367 GDRLLIVANSDSDVKRLKGSREPINNLNENMAVLAGIASVDMLVPFTEDTPQRLIGIILP 426  
 QY 427 DLLVKGDDYKEERIASKEVWANGSGVLYLNFEDGSGTTNIIKKIQ 472  
 DB 427 DLLVKGDDYKEERIASKEVWANGSGVLYLNFENGSITTIVIKIK 472

Search completed: March 19, 2006, 06:06:28  
 Job time : 234 secs



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

# OM protein - protein search, using sw model

Run on: March 18, 2006, 06:06:43 ; Search time 48 Seconds  
(without alignments) 821.590 Million cell updates/sec

Title: US-10-520-820-13

Perfect score: 2393  
Sequence: 1 MKTTLPEFRAAGVWVGDVW.....PDEGCTTIIKKIQDQK 477

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /cgn2\_6/ptcdat/1/1aa/5\_COMB.pep:\*
- 2: /cgn2\_6/ptcdat/1/1aa/6\_COMB.pep:\*
- 3: /cgn2\_6/ptcdat/1/1aa/H\_COMB.pep:\*
- 4: /cgn2\_6/ptcdat/1/1aa/PCrus\_COMB.pep:\*
- 5: /cgn2\_6/ptcdat/1/1aa/RB\_COMB.pep:\*
- 6: /cgn2\_6/ptcdat/1/1aa/backfilltest.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2393	100.0	477	US-09-492-709A-325	Sequence 325, App
2	2227	93.1	499	US-09-489-039A-9208	Sequence 9208, Ap
3	2007	83.9	491	US-09-543-681A-6634	Sequence 6634, Ap
4	1351.5	56.5	526	US-09-252-991A-32984	Sequence 32984, A
5	610.5	25.5	498	US-09-477-962-114	Sequence 114, App
6	197	8.2	342	US-09-489-039A-9668	Sequence 9668, Ap
7	196.5	8.2	503	US-09-252-991A-31882	Sequence 31882, A
8	190.5	8.0	529	US-09-134-000C-4524	Sequence 4524, Ap
9	183	7.6	311	US-09-543-681A-7498	Sequence 7498, Ap
10	171	7.1	356	US-09-489-039A-12757	Sequence 12757, A
11	169	7.1	423	US-09-668-262A-14	Sequence 14, App1
12	169	7.1	423	US-10-427-442-14	Sequence 14, App1
13	167.5	7.0	665	US-09-489-039A-8800	Sequence 8800, Ap
14	165.5	6.9	315	US-09-134-001C-4521	Sequence 4521, Ap
15	165	6.9	330	US-09-489-039A-14043	Sequence 14043, A
16	163.5	6.8	421	US-09-668-262A-16	Sequence 16, App1
17	163.5	6.8	421	US-10-427-442-16	Sequence 16, App1
18	155	6.5	320	US-09-543-681A-5983	Sequence 5983, Ap
19	154.5	6.3	302	US-09-710-279-3062	Sequence 3062, Ap
20	150	6.3	137	US-09-134-001C-5560	Sequence 5560, Ap
21	146.5	6.1	269	US-09-270-767-66416	Sequence 66416, A
22	146	6.1	208	US-09-270-767-61995	Sequence 61995, A
23	143	6.0	293	US-09-668-262A-12	Sequence 12, App1
24	143	6.0	293	US-10-427-442-12	Sequence 12, App1
25	143	6.0	311	US-09-543-681A-5777	Sequence 5777, Ap
26	141	5.9	313	US-09-489-039A-12895	Sequence 12895, A
27	141	5.9	355	US-09-248-796A-16878	Sequence 16878, A

28	139.5	5.8	331	2	US-09-902-540-10221	Sequence 10221, A
29	137.5	5.7	312	2	US-09-134-001C-5583	Sequence 5583, Ap
30	137.5	5.7	347	2	US-08-826-611-2	Sequence 2, App1
31	134.5	5.6	3724	1	US-08-804-227C-10	Sequence 10, App1
32	134.5	5.6	3724	1	US-08-804-198-4	Sequence 4, App1
33	133.5	5.6	308	2	US-08-913-816C-17	Sequence 17, App1
34	131	5.5	143	2	US-09-134-000C-4544	Sequence 4544, Ap
35	130	5.4	859	2	US-09-199-637A-281	Sequence 281, App
36	127	5.3	859	2	US-09-252-991A-21413	Sequence 21413, A
37	125	5.2	314	2	US-09-489-039A-10377	Sequence 10377, A
38	125	5.2	331	2	US-09-543-681A-7111	Sequence 7111, Ap
39	123.5	5.2	328	2	US-09-489-039A-7290	Sequence 7290, Ap
40	121	5.1	301	2	US-09-252-991A-22565	Sequence 22565, A
41	121	5.1	406	2	US-09-543-681A-8124	Sequence 8124, Ap
42	120.5	5.0	269	2	US-09-134-000C-5214	Sequence 5214, Ap
43	119.5	5.0	411	2	US-09-949-016-8628	Sequence 8628, Ap
44	118	4.9	595	2	US-09-902-540-16334	Sequence 16334, A
45	118	4.9	719	2	US-09-902-540-11526	Sequence 11526, A

## ALIGNMENTS

RESULT 1  
US-09-492-709A-325  
Sequence 325, Application US/09492709A  
Patent No. 6720139  
GENERAL INFORMATION:  
APPLICANT: Zykend, Judith  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Trawick, John  
APPLICANT: Forsyth, R. Allyn  
APPLICANT: Friele, Jamie M.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN  
FILE REFERENCE: ELITRA.001A  
CURRENT APPLICATION NUMBER: US/09/492.709A  
CURRENT FILING DATE: 2000-01-27  
NUMBER OF SEQ ID NOS: 485  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 325  
LENGTH: 477  
TYPE: PRT  
ORGANISM: E. Coli  
US-09-492-709A-325

Query Match 100.0%; Score 2393; DB 2; Length 477;  
Best Local Similarity 100.0%; Pred. No. 5.8e-232;  
Matches 477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MKTTLPEFRAAGVWVGDVWLDKRYWGPTSRISPEAPVPVKKVNTIEERGAANYAMNT	60
DB	1	MKTTLPEFRAAGVWVGDVWLDKRYWGPTSRISPEAPVPVKKVNTIEERGAANYAMNT	60
QY	61	ASIGANARLVGLTGIDDAARLSKSLADVNVKCDPVSPHTPTTKRLVLSRNOQLRLD	120
DB	61	ASIGANARLVGLTGIDDAARLSKSLADVNVKCDPVSPHTPTTKRLVLSRNOQLRLD	120
QY	121	FREGFEGVDQPLHERINQSLISGLVLSDYAKGALASVQOMIQLARKGVPLLDPKG	180
DB	121	FREGFEGVDQPLHERINQSLISGLVLSDYAKGALASVQOMIQLARKGVPLLDPKG	180
QY	181	TDPERYRGATLLTPNLSFEFAVVGKCTBEIEYERGMKLADYELSLALVTRSEOGMSLL	240
DB	181	TDPERYRGATLLTPNLSFEFAVVGKCTBEIEYERGMKLADYELSLALVTRSEOGMSLL	240
QY	241	OPKAPLHMPTOAOBYVDVTGAGDTVIYGLAATLAAGNSLEBAQCFANAAAGVVGKLG	300
DB	241	OPKAPLHMPTOAOBYVDVTGAGDTVIYGLAATLAAGNSLEBAQCFANAAAGVVGKLG	300

Qy 301 STVSPTELENAVRGRADTGFVMTTEBELKLAVAARAKGKRVMTNGVFDILHAGHSYL 360  
Db 301 STVSPTELENAVRGRADTGFVMTTEBELKLAVAARAKGKRVMTNGVFDILHAGHSYL 360  
Qy 361 ANARKLDRILVAVNSDASTRLKGDSPVPVPLEQRMIVLGALEAVDWMVVSFEEDTPQRL 420  
Db 361 ANARKLDRILVAVNSDASTRLKGDSPVPVPLEQRMIVLGALEAVDWMVVSFEEDTPQRL 420  
Qy 421 IAGILPDLVKGDDYKPEEIKSGKVMANGGEVVLNPFEDCSTTNIKKIQDQK 477  
Db 421 IAGILPDLVKGDDYKPEEIKSGKVMANGGEVVLNPFEDCSTTNIKKIQDQK 477

## RESULT 2

US-09-489-9208  
Sequence 9208, Application US/09489039A  
Patent No. 6610836  
GENERAL INFORMATION:  
APPLICANT: Gary Breton et. al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
FILE REFERENCE: 2709, 2004001  
CURRENT APPLICATION NUMBER: US/09/489,039A  
CURRENT FILING DATE: 2000-01-27  
PRIOR APPLICATION NUMBER: US 60/117,747  
PRIOR FILING DATE: 1999-01-29  
NUMBER OF SEQ ID NOS: 14342  
SEQ ID NO 9208  
LENGTH: 499  
TYPE: PRT  
ORGANISM: Klebsiella pneumoniae  
US-09-489-9208

Query Match 93.1%; Score 2227; DB 2; Length 499;  
Best Local Similarity 92.2%; Pred. No. 3.3e-215;  
Matches 437; Conservative 25; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MKVTLPEERAGVWVVDVMDLDRYWGPTSRISPEAPVVPVKVNTIEERPGGANVAMNI 60  
Db 23 MKVTLPEERAGVWVVDVMDLDRYWGPTSRISPEAPVVPVKVNTIEERPGGANVAMNI 82  
Qy 61 ASLGANARLVGLTGIDDAARLSKSLADVNVKCDPVSPVPTITTKLRVLSRNOQLRLD 120  
Db 83 ASLGANARLVGLTGIDDAARLSKSLADVNVKCDPVSPVPTITTKLRVLSRNOQLRLD 142  
Qy 121 PEEGSEGVDPQLHERINQALSSIGALVLSYKAGALASVQOMQLARKAGVPLIDPKG 180  
Db 143 PEEGSEGVDPQLHERINQALSSIGALVLSYKAGALASVQOMQLARKAGVPLIDPKG 202  
Qy 181 TDFERYGATLTPNLSEFEAVVKGCKTEBEIYERGMKLIADVELSALLVTRSEQMSL 240  
Db 203 TDFERYGATLTPNLSEFEAVVKGCKTEBEIYERGMKLIADVELSALLVTRSEQMSL 262  
Qy 241 QPGKAPLHPTQAEVVDVYTGADTVIGVLAATLAAGNSLEBAQCFANAAAGVVGKLG 300  
Db 263 QPGKAPLHPTQAEVVDVYTGADTVIGVLAATLAAGNSLEBAQCFANAAAGVVGKLG 322  
Qy 301 STVSPTELENAVRGRADTGFVMTTEBELKLAVAARAKGKRVMTNGVFDILHAGHSYL 360  
Db 323 STVSPTELENAVRGRADTGFVMTTEBELKLAVAARAKGKRVMTNGVFDILHAGHSYL 382  
Qy 361 ANARKLDRILVAVNSDASTRLKGDSPVPVPLEQRMIVLGALEAVDWMVVSFEEDTPQRL 420  
Db 383 ANARKLDRILVAVNSDASTRLKGDSPVPVPLEQRMIVLGALEAVDWMVVSFEEDTPQRL 442  
Qy 421 IAGILPDLVKGDDYKPEEIKSGKVMANGGEVVLNPFEDCSTTNIKKIQD 474  
Db 443 IAGILPDLVKGDDYKPEEIKSGKVMANGGEVVLNPFEDCSTTNIKKIQD 496

## RESULT 3

US-09-543-681A-6634  
Sequence 6634, Application US/09543681A

Patent No. 6605709  
GENERAL INFORMATION:  
APPLICANT: GARY BRETON  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
FILE REFERENCE: 2709, 1002-001  
CURRENT APPLICATION NUMBER: US/09/543,681A  
CURRENT FILING DATE: 2000-04-05  
PRIOR APPLICATION NUMBER: US 60/128,706  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 8344  
SEQ ID NO 6634  
LENGTH: 491  
TYPE: PRT  
ORGANISM: Proteus mirabilis  
US-09-543-681A-6634

Query Match 83.9%; Score 2007; DB 2; Length 491;  
Best Local Similarity 82.2%; Pred. No. 4.6e-193;  
Matches 389; Conservative 43; Mismatches 41; Indels 0; Gaps 0;

Qy 1 MKVTLPEERAGVWVVDVMDLDRYWGPTSRISPEAPVVPVKVNTIEERPGGANVAMNI 60  
Db 18 MKVTLPEERAGVWVVDVMDLDRYWGPTSRISPEAPVVPVKVNTIEERPGGANVAMNI 77  
Qy 61 ASLGANARLVGLTGIDDAARLSKSLADVNVKCDPVSPVPTITTKLRVLSRNOQLRLD 120  
Db 78 ASLGANARLVGLTGIDDAARLSKSLADVNVKCDPVSPVPTITTKLRVLSRNOQLRLD 137  
Qy 121 PEEGSEGVDPQLHERINQALSSIGALVLSYKAGALASVQOMQLARKAGVPLIDPKG 180  
Db 138 PEEGSEGVDPQLHERINQALSSIGALVLSYKAGALASVQOMQLARKAGVPLIDPKG 197  
Qy 181 TDFERYGATLTPNLSEFEAVVKGCKTEBEIYERGMKLIADVELSALLVTRSEQMSL 240  
Db 198 TDFERYGATLTPNLSEFEAVVKGCKTEBEIYERGMKLIADVELSALLVTRSEQMSL 257  
Qy 241 QPGKAPLHPTQAEVVDVYTGADTVIGVLAATLAAGNSLEBAQCFANAAAGVVGKLG 300  
Db 258 QPGKAPLHPTQAEVVDVYTGADTVIGVLAATLAAGNSLEBAQCFANAAAGVVGKLG 317  
Qy 301 STVSPTELENAVRGRADTGFVMTTEBELKLAVAARAKGKRVMTNGVFDILHAGHSYL 360  
Db 318 STVSPTELENAVRGRADTGFVMTTEBELKLAVAARAKGKRVMTNGVFDILHAGHSYL 377  
Qy 361 ANARKLDRILVAVNSDASTRLKGDSPVPVPLEQRMIVLGALEAVDWMVVSFEEDTPQRL 420  
Db 378 ANARKLDRILVAVNSDASTRLKGDSPVPVPLEQRMIVLGALEAVDWMVVSFEEDTPQRL 437  
Qy 421 IAGILPDLVKGDDYKPEEIKSGKVMANGGEVVLNPFEDCSTTNIKKIQ 473  
Db 438 IAGILPDLVKGDDYKPEEIKSGKVMANGGEVVLNPFEDCSTTNIKKIQ 490

## RESULT 4

US-09-252-991A-32984  
Sequence 32984, Application US/09252991A

Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196, 136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 32984  
LENGTH: 526  
TYPE: PRT

ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-32984

Query Match 56.5%; Score 1351.5; DB 2; Length 526;  
Best Local Similarity 57.9%; Pred. No. 5.4e-127;  
Matches 274; Conservative 70; Mismatches 128; Indels 1; Gaps 1;

QY 1 MKTLPFEFAGVWVVDVLDLRYVGPTRISPEAVPVVKNVTIERPGGAANVAMNT 60  
DB 54 MKLSMPFDDAPVTVVDVMDLDRVMGATSRISPEAVPVVQHQHDPGGAANVAMNT 113  
QY 61 ASIGANRLVGLTGTIDDAABALSKSLADVNVKCFVSPVPTHTTKLRYLSRNQOLRLD 120  
DB 114 ALGAGQALLVGVTRDGLDLSLNSLTLAAGVDAFQIDISOPTIVKLKRVSRHQQLRVD 173  
QY 121 FEEGFEQVDPQLHERINQALSSIGALVLSDYAKGALASVQOMIQLARKAGVPLIDPKG 180  
DB 174 FEEFPR-TDAALAVDVESLLAKVVLVLSYKGLQNHQVLIQAARANIPLVADPKG 232  
QY 181 TDERRRGATLLTPNLSEFPAVVGKCTEERIVERGKCLADYELSLVLTTRSEQMSLL 240  
DB 233 KDAIVKGAELTPNLSEFPTIVGRCADEALVAKGALMSLDLGLVTRGHEGHTLL 292  
QY 241 OPGKAPLHPTQAEVYDTGAGPTVIGVLAATLAAGNSLEBAQCFPANAAGVVGKLG 300  
DB 293 RDGPALHLPARAREVDVTGAGDTVSTLAAALAAEELPSAVGLANTLAIYVGKLG 352  
QY 301 STVSPLELENVGRADTGFVMTSEELKLAVAARKRGKVNTGVPDILAHGVSYL 360  
DB 353 AALSAPELRABVQEGSBRGVGLDGLQLLAIEDRAHGEKIVTNGCPDILAHGVSYL 412  
QY 361 ANARKLDRLIVANSASDSTRKLGDSRPVNPLEQRITVGALEAVDVMVSPBEDTQRL 420  
DB 413 EOAAQODRLIVGNDASVTRKGVGRPNISVDRRAVLAGLADVMVSPBEDTQRL 472  
QY 421 IAGILPDLVKGDDYKPEBIAGSKVANGGEVVLNFEQDCSTNTIITKIQ 473  
DB 473 LEQVRPVLVKGDDYGVQVGAQIVKAYGGEVAVLVLVENSITTAIVETKIQ 525

## RESULT 5

US-09-477-962-114  
Sequence 114, Application US/09477962  
Patent No. 6927286  
GENERAL INFORMATION:  
APPLICANT: SHEN, BEN  
APPLICANT: DU, LIANGCHENG  
APPLICANT: SANCHEZ, CESAR  
APPLICANT: CHEN, MEI  
APPLICANT: EDWARDS, DANIEL J.  
TITLE OF INVENTION: BLEDOMYCIN GENE CLUSTER COMPONENTS AND THEIR USES  
FILE REFERENCE: 4077-895820US  
CURRENT APPLICATION NUMBER: US/09/477,962  
CURRENT FILING DATE: 2000-01-05  
PRIOR APPLICATION NUMBER: 60/115,435  
PRIOR FILING DATE: 1999-01-06  
PRIOR APPLICATION NUMBER: 60/118,848  
PRIOR FILING DATE: 1999-02-05  
NUMBER OF SEQ ID NOS: 133  
SOFTWARE: PatentIn Ver. 3.0  
SEQ ID NO 114  
LENGTH: 498  
TYPE: PRT  
ORGANISM: Streptomyces verticillius  
FEATURE:  
OTHER INFORMATION: ORF9  
US-09-477-962-114

Query Match 25.5%; Score 610.5; DB 2; Length 498;  
Best Local Similarity 37.7%; Pred. No. 2.1e-52;  
Matches 182; Conservative 68; Mismatches 196; Indels 37; Gaps 15;  
QY 13 VMVGVMDLBRVWYGPTSRISPEAVPVVKNVTIERPGGAANVAMNTASIGANRLVGL 72

DB 17 VLVIGVYIIDTYWGAITSGLCRSPVPATVLTISVAGCGGANVAVNLTALSGEPVLSA 76  
QY 73 TGIDDAARALSKSL--ADVNVKCDPVSPVPTHTTKLRYLSRNQOLRLDFEEGFEQVDP 130  
DB 77 TGDDBRGRRLRLALRARDVDTGGLFVQ-PGRVTIVGRVADQOMLRLD--EG--GEHP 131

QY 131 QP-----LHERINQALSSIGALVLSDYAKGALASVQOMIQLA--RKAQVVL--DE 178  
DB 132 LPATDTGSLRLRAAGLLPADAVALVSDGYG-VNEPDTVAALAHREIGPSTLVDSR 190  
QY 179 KGTDFERYGATLLTPNLSEF-----BAVVGKCTEERIVERGKCLADYELSLVLT 231  
DB 191 RPARFTALR-ASAVKNNHAAALRLDAGERPPPPAA--DMAALGRLRLTGAEEVALT 248  
QY 232 RSEQMSLLQPKAPLHPTQAE--VYDTGAGDTVIGVLAATLAAGNSLEBAQCFPANA 289  
DB 249 LDADGSLTFERDRPVRTFARGSRAPVTAAGADAFPTAALTALAAAGDASAAAEIASA 308  
QY 290 AAGVVGKLGTSVSPLELENVGRADTGFVMTSE-ELKLAVAARKRGKVNTGVP 348  
DB 309 AAGTAAVAPGTSTWMADELRLRLG--TGKVCRTGTLPARLIDPARDR--RVVETNGC 363  
QY 349 FDLIHAHSVYLANARKLDRLIVANSASDSTRKLGDSRPVNPLEQRITVGALEAVDVM 408  
DB 364 FDLIHAHSVYLANARKLDRLIVANSASDSTRKLGDSRPVNPLEQRITVGALEAVDVM 423  
QY 409 VVSPBEDTQRLIAGILPDLVKGDDYKPEBIAGSKVANGGEVVLNFEQDCSTNTI 468  
DB 424 VVFPDDSPALILALRPVYAKGDDYTLATLEAPLVORLGGVVLHLSVADTSTTDII 483  
QY 469 KKI 471  
DB 484 RRI 486

## RESULT 6

US-09-489-039A-9668  
Sequence 9668, Application US/09489039A  
Patent No. 6610836  
GENERAL INFORMATION:  
APPLICANT: Gary Breton et. al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 2709.2004001  
CURRENT APPLICATION NUMBER: US/09/489,039A  
CURRENT FILING DATE: 2000-01-27  
PRIOR APPLICATION NUMBER: US 60/117,747  
PRIOR FILING DATE: 1999-01-29  
NUMBER OF SEQ ID NOS: 14342  
SEQ ID NO 9668  
LENGTH: 342  
TYPE: PRT  
ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-9668

Query Match 8.2%; Score 197; DB 2; Length 342;  
Best Local Similarity 26.9%; Pred. No. 5.1e-11;  
Matches 90; Conservative 52; Mismatches 162; Indels 30; Gaps 14;

QY 2 KYTLPEFERAGV--WVGVV-MLDRYWYGPTSRISPEAVPV--VKNVTIERPGGA 54  
DB 17 RAALP---RAGIERMMSGKVCVFGSFNFDMAKVD-RFPVPEBSLVACGSMTSAGKGA 72  
QY 55 NVAMNTASIGANRLVGLTGTIDDAABALSKSLADVNVKCDPVSPVPTHTTKLRYLSRN 113  
DB 73 NQATLAKGAVVHYIGKIGNDTFGFAHRHLKGVGFNAVTLVVAETPTGALIVAGN 132  
QY 114 --QQLRLDFEFGFEQVDPQLHERINQALSSIG-ALVLSDYAKGALASVQOMIQLARKA 170  
DB 133 DARNMLAVD-----PANNTVTVDDERLACGCPALGCAADVVLVQLENNISALBOYIDGKA 187  
QY 171 GVPVLIDP---KGTDFERYGATLLTPNLSEFPAVVGKCTEERIVERGKCLADYELSA 227

Db 188 GALVILNAPQVPEHALLRKVDLTTPVATEAGMTGRRVDSLTPAAEAADVLAQAGRN 247  
Qy 228 LLYVTRSEGMGLLPQ-KAPLHMPQAOEVVDYVAGDTVIGVLAATLAAGNSLEACFP 286  
Db 248 VITLGASGALLSHRGVSP1-PCPSPHPTDTTGAGFAGALAAARLACGEP1QAAARF 305  
Qy 287 AANAAGVVGKLTGTSTVSP1EL---ENAVRGRAD 317  
Db 306 AAAYAAVSVEKQAGASSL-PEVLEAQRRLRAAD 338

## RESULT 7

US-09-252-991A-31882  
; Sequence 31882, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252.991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 31882  
; LENGTH: 503  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-31882

Query Match 8.2%; Score 196.5; DB 2; Length 503;  
Best Local Similarity 26.4%; Pred. No. 1.1e-10;

Matches 88; Conservative 47; Mismatches 163; Indels 35; Gaps 11;

Qy 6 PEFER-----AGVVVGVGDMIDRYWYPTSRISPEAPVVPVKVNTIERPGG-AAN 55  
Db 184 PSLERGNKDKATMQAKVIVGSLNMDLVYRAPRLPRGE---TLAGSFTTITGGKGAN 239  
Qy 56 VAMNIASIGANARLVLTGIDDAARALSKSLADVNVKDFVS-VPHETITKLAVLSRQ 114  
Db 240 QAVAAARLGAIVAMIGCLGDAYGDOQLYRALQAEIGIDQGERVAGESSVALIVVDDSS 299  
Qy 115 QLILDFEEREGVDPOPL-HERINQALSSIGALVSDYAKGALASVQOMIQLARKAGV 172  
Db 300 QMAIVTVAGSGHLSPAVLARHEHLLEQ---AQVWCQLESPLFETGHTLRRAHALGK 354  
Qy 173 PVLIDP-----KGTDFERYRGATLLTPNLSFEFAVVGKCTEETIVE--RGMKLIDVYL 225  
Db 355 TVIINPAPATRDVPADMLPLVDYLVFNTESEL---CRLPVDSLESGRAPAEIRRENGA 411  
Qy 226 SALLVTRSEQMSLLQPGKAPLHMPQAOEVVDYVAGDTVIGVLAATLAAGNSLEACF 285  
Db 412 GRVIVTIGAGGALLVGEGRVE-HFPVARVKALDTTAAGDTFVGGPAAALAGLDEAAAIR 470  
Qy 286 PANAAGVVGKLTGTSTVSP1--ELENAVGRGA 316  
Db 471 FGQAAAISVTRLAGQTSIPREVERALGEEA 503

## RESULT 8

US-09-134-000C-4524  
; Sequence 4524, Application US/09134000C  
; Patent No. 6617156  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 032796-032  
; CURRENT APPLICATION NUMBER: US/09/134.000C

; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/055,778  
; PRIOR FILING DATE: 1997-08-15  
; NUMBER OF SEQ ID NOS: 6812  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4524  
; LENGTH: 529  
; TYPE: PRT  
; ORGANISM: Enterococcus faecalis  
US-09-134-000C-4524

Query Match 8.0%; Score 190.5; DB 2; Length 529;  
Best Local Similarity 24.5%; Pred. No. 4.8e-10;

Matches 81; Conservative 53; Mismatches 126; Indels 71; Gaps 11;

Qy 13 VNVVGVDMIDRYWY-----GPTSRISPEAPVVPVKVNTIERPGGAANVAMNIASIGA 65  
Db 219 MIVVGSINVDNIVYSTNLPHNGKTNPLSYAKFP-----GGKGLNQAAGLTYLGH 268  
Qy 66 NARLVGLTGIDDAARALSKSLADVNVKDFVSVPHTITKLRLSRNQLIRLD----- 120  
Db 269 QATLIGCLSDTPANTLYKELKRYHTTD-----GTRIDTERGQAVIVETSGDS 320  
Qy 121 -----FEEGFEVDPOPLHERINQALSSIGALVSDYAKGALASVQOMIQLARKAGVVL 176  
Db 321 MISILPGANTALTPKIAQCKHLPMDASFCLIQTEI---PLSAVERKACEITQSHGVPIIL 377  
Qy 177 DP-----KGTDFERYRGATLLTPNLSFEFAVVGKCTEETIVE--IVERGMGLIA 221  
Db 378 KPAAIHIPVNIIEKDF-----FVNEDELLEOPTGTGLLEEAAVFLKGVGVNV- 428  
Qy 222 DYERISALLVTRSEQMSLLQPGKAPLHMPQAOEVVDYVAGDTVIGVLAATLAAGNSLE 281  
Db 429 -----IVTLGKGV-LKTPQVCHYPATENTIAVDSGASPSFALSALYSKGYPT 480  
Qy 282 EACFPANAAGVVGKLTGT--STVSP1ELEN 310  
Db 481 AAIOIAIOAAGFSVSKGVIDSIVDHTLEN 511

## RESULT 9

US-09-543-681A-7498  
; Sequence 7498, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BRETTON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
; FILE REFERENCE: 2709.1002-001  
; CURRENT APPLICATION NUMBER: US/09/543.681A  
; CURRENT FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 7498  
; LENGTH: 311  
; TYPE: PRT  
; ORGANISM: Proteus mirabilis  
US-09-543-681A-7498

Query Match 7.6%; Score 183; DB 2; Length 311;  
Best Local Similarity 24.3%; Pred. No. 1.1e-09;

Matches 63; Conservative 47; Mismatches 141; Indels 8; Gaps 5;

Qy 51 GGAANVAMNIASIGANARLVLTGIDDAARALSKSLADVNVKDFVSVPHTITKLRLV 110  
Db 44 GKGANQVAAAGRGANTFLACLNDDIGSARAQLITDKIDTDCLEL-IDDEATGVALL 102  
Qy 111 SRNQLLR-LDFEEREGVDPOPLHERINQALSSIGALVSDYAKGALASVQOMIQLARK 169  
Db 103 PMAOQGANVIGIHAGANGRLNREVERKYNVITKADALLMO--LESFLDSVLAAQAVAKQ 160  
Qy 170 AGVPVLIDP---KGTDFERYRGATLLTPNLSFEFAVVGKCTEETIVERGMKLIDVYELS 226

Db 161 BSVGVILNPAAPAKLPELLESLVDITPNETETRYLGVAVIDDESQKRSNIHHDGIE 220  
 Qy 227 ALVTSEQSGMSLLQPKGAPLHPTQAEVYDVGADTVIGVLAATLAAGNSLEACFP 286  
 Db 221 TVLITGSRGVWVSQNNQGTW-VPAFKVTAIDTIAAGDTGAFITALLBGLPMMKAIKF 279  
 Qy 287 ANAAGVGVGKLTSTSP 305  
 Db 280 AHAALAAVTRAGAQSPV 298

## RESULT 10

US-09-489-039A-12757  
 ; Sequence 12757, Application US/09489039A  
 ; Patent No. 6610836  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gary Breton et. al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
 ; FILE REFERENCE: 2709.2004001  
 ; CURRENT APPLICATION NUMBER: US/09/489,039A  
 ; PRIOR FILING DATE: 2000-01-27  
 ; PRIOR APPLICATION NUMBER: US 60/117,747  
 ; PRIOR FILING DATE: 1999-01-29  
 ; NUMBER OF SEQ ID NOS: 14342  
 ; SEQ ID NO 12757  
 ; LENGTH: 356  
 ; TYPE: PRT  
 ; ORGANISM: Klebsiella pneumoniae  
 US-09-489-039A-12757

Query Match 7.1%; Score 171; DB 2; Length 356;  
 Best Local Similarity 24.2%; Pred. No. 2.3e-08;

Matches 80; Conservative 40; Mismatches 152; Indels 58; Gaps 11;  
 Qy 28 PTHSISEAPVPVVKVNTIERPGG-AAVAMNINASIGANARIVGLGIDDAARLSKL 86  
 Db 63 PVVAIPRGGVAITE-QIRLNPAGTANAGANITNAKLGITAAVACGDEKADFTILASY 120  
 Qy 87 ADVNVKCDPVV-----VPTHPTITKRLVLSNOCILRLDFEEGEGV-DPOPLH 134  
 Db 121 ARIGDLSLQRTALKETSATILPIRNGERPALHCGASDALFVSAEDAVLDCRFLH 180  
 Qy 135 ERINQALSSIGALVLSYAKALASVQOMIQARKAGPV--LIDPKGDFERYGATL 191  
 Db 181 H-----GGTCLTAAHQGSA-----RLQAAKARGVTTSPDLIAENEETLELRP--- 226  
 Qy 192 LTPULSRF-----BAVVGKCKTEBIEVERGKLIADVELSALLVTRSEQSGMSLLQPKAP 246  
 Db 227 LRPVDYFMPGLEBAFLSGETPEALGRFFLAG---VGTCLIKDGENSGMWLIGRDGP 283  
 Qy 247 LHMPTQAEVYDVGADTVIGVLAATLAAGNSLEACFPANAAAGVGVGKLTSTSP 306  
 Db 284 QHIAFWVEAVDTTGGCDSDYCGSFIALARGLSYKACDVAANVAALVATGMSGDA----- 339  
 Qy 307 ELENAVRGADTGFVGVTSEELKLAVAAR 336  
 Db 340 -----GVVDMEQTQAFMAHR 355

## RESULT 11

US-09-668-262A-14  
 ; Sequence 14, Application US/09668262A  
 ; Patent No. 6596926  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Falco, Carl  
 ; APPLICANT: Ramodu, Omolayo O.  
 ; APPLICANT: Kinney, Tony  
 ; APPLICANT: Rafalski, Antoni  
 ; TITLE OF INVENTION: Phosphatidylocholine Biosynthetic Enzyme  
 ; FILE REFERENCE: BB1403 US NA  
 ; CURRENT APPLICATION NUMBER: US/09/668,262A

; CURRENT FILING DATE: 2000-09-22  
 ; PRIOR APPLICATION NUMBER: 60/155626  
 ; PRIOR FILING DATE: 1999-09-23  
 ; NUMBER OF SEQ ID NOS: 20  
 ; SOFTWARE: Microsoft Office 95  
 ; SEQ ID NO 14  
 ; LENGTH: 423  
 ; TYPE: PRT  
 ; ORGANISM: Oryza sativa  
 US-09-668-262A-14

Query Match 7.1%; Score 169; DB 2; Length 423;  
 Best Local Similarity 30.5%; Pred. No. 4.8e-08;  
 Matches 67; Conservative 24; Mismatches 79; Indels 50; Gaps 9;

Qy 274 LAAGNSLEACFPANAAAGVGVGKLTSTSPS-----PIELENAVRGADTGFVMTSEEL 328  
 Db 1 MEAGAGSSSAKLVAACVIGIV--LGASVVALHLAGPVAILPALRRRRAPRR----- 50  
 Qy 329 KLAVAARRGKGVNTNGVFDLILHAGVSYLANARKDRLIVANSASTRLKGDGR 388  
 Db 51 ----RRRRRPVAVV-DCGFDMMHYGHCNALRQARALDDELIVGVSDHEITANKGP-- 103  
 Qy 389 PVNPLRQRMIVGALRAVDVWVSFEEDTP-----QRLIAGILPDLVYK 431  
 Db 104 PVTPLHERLIM--VAAVKMHVVIDPAPVATIEDFMNKLFNENYNDIYIHGDDPCLLPD 160  
 Qy 432 GGDYKPEELIAGSKVWANGGEVLVLPEDGCTTNIKKI 471  
 Db 161 GTD-----AVLAKKV-----GRFKQIRKTEGVSTTDIVGM 192

## RESULT 12

US-10-427-442-14  
 ; Sequence 14, Application US/10427442  
 ; Patent No. 6960704  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Falco, Carl  
 ; APPLICANT: Ramodu, Omolayo O.  
 ; APPLICANT: Kinney, Tony  
 ; APPLICANT: Rafalski, Antoni  
 ; TITLE OF INVENTION: Phosphatidylocholine Biosynthetic Enzyme  
 ; FILE REFERENCE: BB1403 US NA  
 ; CURRENT APPLICATION NUMBER: US/10/427,442  
 ; CURRENT FILING DATE: 2003-05-01  
 ; PRIOR APPLICATION NUMBER: US/09/668,262A  
 ; PRIOR FILING DATE: 2000-09-22  
 ; PRIOR APPLICATION NUMBER: 60/155626  
 ; PRIOR FILING DATE: 1999-09-23  
 ; NUMBER OF SEQ ID NOS: 20  
 ; SOFTWARE: Microsoft Office 95  
 ; SEQ ID NO 14  
 ; LENGTH: 423  
 ; TYPE: PRT  
 ; ORGANISM: Oryza sativa  
 US-10-427-442-14

Query Match 7.1%; Score 169; DB 2; Length 423;  
 Best Local Similarity 30.5%; Pred. No. 4.8e-08;  
 Matches 67; Conservative 24; Mismatches 79; Indels 50; Gaps 9;

Qy 274 LAAGNSLEACFPANAAAGVGVGKLTSTSPS-----PIELENAVRGADTGFVMTSEEL 328  
 Db 1 MEAGAGSSSAKLVAACVIGIV--LGASVVALHLAGPVAILPALRRRRAPRR----- 50  
 Qy 329 KLAVAARRGKGVNTNGVFDLILHAGVSYLANARKDRLIVANSASTRLKGDGR 388  
 Db 51 ----RRRRRPVAVV-DCGFDMMHYGHCNALRQARALDDELIVGVSDHEITANKGP-- 103  
 Qy 389 PVNPLRQRMIVGALRAVDVWVSFEEDTP-----QRLIAGILPDLVYK 431  
 Db 104 PVTPLHERLIM--VAAVKMHVVIDPAPVATIEDFMNKLFNENYNDIYIHGDDPCLLPD 160

QY 432 GGDYKPEEISKEVWANGEVLTINFDGCSITNIIRKI 471  
 Db 161 GTD----AVALAKKV-----GFFKQIKRTGEGSTTIDYGRM 192

# RESULT 13

US-09-489-039A-8800  
 ; Sequence 8800, Application US/09489039A  
 ; Patent No. 6610836  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gary Breton et. al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
 ; FILE REFERENCE: 2709,2004001  
 ; CURRENT FILING DATE: 1997-11-08  
 ; PRIOR FILING DATE: 1997-08-14  
 ; PRIOR APPLICATION NUMBER: US 60/055,779  
 ; NUMBER OF SEQ ID NOS: 5674  
 ; SEQ ID NO 4521  
 ; LENGTH: 315  
 ; TYPE: PRT  
 ; ORGANISM: Staphylococcus epidermidis  
 ; US-09-134-001C-4521

Query Match 7.0%; Score 167.5; DB 2; Length 665;  
 Best Local Similarity 22.5%; Pred. No. 1.5e-07;  
 Matches 120; Conservative 91; Mismatches 203; Indels 119; Gaps 27;

QY 9 ERAGVWVGDVMDLRYWYGTSPRISPEAPVPPVVKNTIEBPFGGANVANNIATSGANAR 68  
 Db 34 KRLDVICIGRAVAVDLVAQOIGSRLEDAV-----SPAKYIGSSGNGVAFGTALQGLSKA 86  
 QY 69 LVGLTGIDDAARALSKSLADVNVKCDFFVSPHTPTTKRLVLS--RNQQLRLDFEE---G 124  
 Db 87 MLAVGGBEHNKRFLETLNRAGVTEYI--ITDKSRITLALVWLGIKQDETPLIYRDNCA 145  
 QY 125 FEVDPOPLHERINQALSSIGALVLS---DYAKGALSVQOMIQLARKAGV----- 172  
 Db 146 DMALTPDISE---EYIASSPALAVTGTLSHA--NTRAAVAKALEVARRHGLRFLALDIDY 201  
 QY 173 -PVL--IDPKGTDEERRGATLLTPNLSE---FEAVVGKKTTEEELVERG-----M 217  
 Db 202 RPLVMGLSLDGGTFLBSGPVTSQLOEVLHLPDLVVG---TEEBPHIAGSITDLTLAL 258  
 QY 218 KLIDVELSALLVTRSEQMSLQP-----GKAPLHMPTOAQEVYDVYTGADTVIGVLA 271  
 Db 259 KAVNNAATKATITVCRGPMGCVVLEGGDIPDSWDQVPLQOQVAV--EVLNVLAGDAPMGGL 317  
 QY 272 ATLAAGNSLEBACFPANAAAGVVKLGSTVSP--IELENAVGRADTGFVMTTEELK 329  
 Db 318 RGMINDSEWQACRYANACGALVVSRRHCADAMPFTKVELDDYLQ--RAESVPRPVDRLN 376  
 QY 330 LAVAAARERGERKVMNTKVPILAGHVSILAN--ARKLGD-----RLIVANSDAST 380  
 Db 377 HLAHVTSRRQOMPELCIFAFD-----HKKQLADLARETGRDEACIPOLKLLLAABAAA 431  
 QY 381 KRLKDSRPVNPLEORMIVLG-----ALBAVD---VSPFEEDTPORLIAGILP 426  
 Db 432 QEA-----GLDORSGIADGTGYGQALALNATGKGMWIRPFIELBSRRLRLEHGNIG 483  
 QY 427 DLAVKGD-----YKPEIAGSK-----EVM-----ANGGEVLV 455  
 Db 484 SOLI---DMPLEHVVKLVFYHPDDPALRAEQDALLLEVVQAQCKSGHELL 533

# RESULT 14

US-09-134-001C-4521  
 ; Sequence 4521, Application US/09134001C  
 ; Patent No. 6380370  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn Doucette-Stamm et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: GTC-007  
 ; CURRENT APPLICATION NUMBER: US/09/134,001C  
 ; CURRENT FILING DATE: 1998-08-13  
 ; PRIOR FILING DATE: 1997-11-08  
 ; PRIOR APPLICATION NUMBER: US 60/064,964  
 ; PRIOR FILING DATE: 1997-11-08  
 ; PRIOR APPLICATION NUMBER: US 60/055,779  
 ; PRIOR FILING DATE: 1997-08-14  
 ; NUMBER OF SEQ ID NOS: 5674  
 ; SEQ ID NO 4521  
 ; LENGTH: 315  
 ; TYPE: PRT  
 ; ORGANISM: Staphylococcus epidermidis  
 ; US-09-134-001C-4521

Query Match 6.9%; Score 165.5; DB 2; Length 315;  
 Best Local Similarity 22.6%; Pred. No. 6.6e-08;  
 Matches 77; Conservative 56; Mismatches 142; Indels 65; Gaps 13;

QY 2 KVTLPFERAGVWVGDVMDLRYWYGTSPRISPEAPVPPVVKNTIEBPFG--AANVANN 59  
 Db 3 KVEVEVNNKKVIVISTVNDKFL--NVKRPK--FGELHINQAKFEGGKGANQAL 58  
 QY 60 IASLGANRLVGLTGIDDAARALSKSLADVNVKCDFFVSPHTPTTKRLVLSRNQQLRL 119  
 Db 59 ASRLADDTTFISRYGKDGANNFTLE-----DEKKAQIH-----TOYILTS 98  
 QY 120 DFEF---GFEVDPOPL-----LHERINQALS-----SIGALVSDYAKAL-----ASVQ 161  
 Db 99 ESEETQOAFITVDEAQNITLVYGANNITLSADVMSVDAFIGADPVVAQLVPEARE 158  
 QY 162 QMIQLARKAGVPLIDP-----KQTDPERRGATLLTPNLSEFEAVVGKCTEE 210  
 Db 159 QAKIRAKQKITVLLNPAPALBELPKSLBLTLD-----IIPNTEAELLTGISINNE 210  
 QY 211 EYERGMKLIADVELSALLVTRSEQMSLQPKAPLHMPTOAQEVYDVYTGADTVIGVLA 270  
 Db 211 SDMKETATYFLDGIISAVLITLGEQGYTCAYOBYKM--IPACNVKAIITTAAGDTFIGAF 269  
 QY 271 AATLAAG-NSLEBACFPANAAAGVVKLGSTVSPLEIELE 309  
 Db 270 LSELINDSLNLESAIRLANQASSLTVQRKGAQASIFTRKE 309

RESULT 15  
 US-09-489-039A-14043  
 ; Sequence 14043, Application US/09489039A  
 ; Patent No. 6610836  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gary Breton et. al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
 ; FILE REFERENCE: 2709,2004001  
 ; CURRENT FILING DATE: 1997-11-08  
 ; PRIOR FILING DATE: 1997-08-14  
 ; PRIOR APPLICATION NUMBER: US 60/055,779  
 ; NUMBER OF SEQ ID NOS: 14342  
 ; SEQ ID NO 14043  
 ; LENGTH: 330  
 ; TYPE: PRT  
 ; ORGANISM: Klebsiella pneumoniae  
 ; US-09-489-039A-14043

Query Match 6.9%; Score 165; DB 2; Length 330;  
 Best Local Similarity 23.7%; Pred. No. 8e-08;  
 Matches 65; Conservative 44; Mismatches 135; Indels 30; Gaps 6;  
 QY 51 GGANVANNIATSGANARLVGLTGIDDAARALSKSLADVNVKCDFFVSPHTPTTKRLV 110  
 Db 63 GKGANQAVNAGRGADIACTGDDDIGERRIRQLASDKI-----DVAPYRAV 111  
 QY 111 SRNQLRLDFEGFEGVDPOPLHERINQALS-----SIGALVSDYAKALA 158

```

Db      112 AGEATGVALITVNA-BGENVIGIHAGANAALSVQVEAEKERIASAQLIMQ--LESPLK 168
Qy      159 SVQOMIOLAKKAGVPLIDP---KGTDFERYRGATLLTPNLSEFEAVVGCKTEEEIVER 215
Db      169 SVIAAKIAHHHTTTLNPPAPARELPDELLALVDIITPNETEAKLTGIRVEDDAK 228
Qy      216 GKKLIADYEUSALIVTRSEQMSLLQPGKAPLHMPTOAQEVYDVTGAGDTVIGVLAATLA 275
Db      229 AADVLAAKGIGTWTLLSGRGVWLSAGGS-RRIPGRVQALDTIAAGDTFNGALVTALL 287
Qy      276 AGNSLEACFPANAAGVVGKLGSTVSPLE 309
Db      288 EGTALPEAIRPAHAAAIAVTRKGAQPSVPWRTS 321

```

Search completed: March 18, 2006, 06:08:07  
 Job time : 50 secs

**THIS PAGE LEFT BLANK**



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using bw model

Run on: March 18, 2006, 19:03:29 ; Search time 7122 Seconds  
(without alignments)  
11445.312 Million cell updates/sec

Title: US-10-520-820-28

Perfect score: 1434  
Sequence: 1 atgaagaatcagctgcgcaga.....aacagataaaaaagctaa 1434

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5083141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

GenSmb1:\*  
1: gb\_ba:\*  
2: gb\_in:\*  
3: gb\_env:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pac:\*  
7: gb\_ph:\*  
8: gb\_pr:\*  
9: gb\_ro:\*  
10: gb\_rst:\*  
11: gb\_ey:\*  
12: gb\_un:\*  
13: gb\_vl:\*  
14: gb\_hlg:\*  
15: gb\_pl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	1434	100.0	1434 6	BD268493 Gene iden
2	1434	100.0	1434 6	CQ779270 Sequence
3	1434	100.0	1434 6	AR493607 Sequence
4	1434	100.0	1434 6	AX189075 Sequence
5	1434	100.0	1434 6	AX363670 Sequence
6	1434	100.0	25638 1	ECU28379 Sequence
7	1434	96.8	110000 1	U00096.31
8	1387.6	96.8	110000 1	AE005174_39
9	1387.6	96.8	110000 1	AE005174_40
10	1387.6	96.3	300359 1	BA000007_39
11	1381.2	95.9	292906 1	AE016767 Escherich
12	1373.2	95.8	110000 1	AE005674_31
13	1373.2	70.6	1500 6	AR385308 Sequence
14	1011.8	70.3	258050 1	AL627278 Salmone11
15	1008.6	70.1	1431 1	AE016844 Salmone11
16	1005.4	70.0	20604 1	AE008847 Salmone11
17	1003.6			
18	1003.6			

C 19	1003.8	70.0	110000 1	AE017220_33	Continuation (34 o
C 20	1002.2	69.9	110000 1	CP000026_31	Continuation (32 o
C 21	999	69.7	1476 1	AF163661	AF163662 Salmone11
C 22	969.8	67.6	1464 1	AF163662	Continuation (41 o
C 23	880.8	61.4	110000 1	BX936398_40	AE013955 Yersinia
C 24	879.2	61.3	10641 1	AE013955	Continuation (41 o
C 25	879.2	61.3	208050 1	AJ414144	Continuation (41 o
C 26	879.2	61.3	290924 1	AE017138	Continuation (41 o
C 27	873.6	60.9	110000 1	BX950851_40	Continuation (41 o
C 28	870.6	60.7	1353 6	AR385378	Continuation (41 o
C 29	784.8	54.7	348498 1	BX571872	Continuation (41 o
C 30	784.8	54.7	349980 6	AX770910	Continuation (41 o
C 31	724.2	50.5	1476 6	AR377456	Continuation (41 o
C 32	668	46.6	10373 1	AE004313	Continuation (41 o
C 33	645.2	45.0	1828 1	AY127572	Continuation (41 o
C 34	635.2	44.3	1465 1	U32828	Continuation (41 o
C 35	635.2	44.3	110000 6	BD426631_15	Continuation (41 o
C 36	635.2	44.3	110000 6	AR274513_15	Continuation (41 o
C 37	635.2	44.3	110000 6	AR632719_15	Continuation (41 o
C 38	628.8	43.8	110000 1	BA000037_05	Continuation (41 o
C 39	626.6	43.7	110000 1	BA000031_04	Continuation (41 o
C 40	625.6	43.6	110000 1	CP000057_14	Continuation (41 o
C 41	625.6	43.6	349980 6	CQ873160	Continuation (41 o
C 42	624	43.5	301235 1	AE016799	Continuation (41 o
C 43	622.4	43.4	110000 1	AE016827_12	Continuation (41 o
C 44	622	43.4	347994 1	CR378664	Continuation (41 o
C 45	617.4	43.1	10747 1	AE006127	Continuation (41 o

#### ALIGNMENTS

RESULT 1	BD268493	1434 bp	DNA	linear	PAT 17-JUN-2003
LOCUS	BD268493				
DEFINITION	Gene identified as being necessary in growth of Escherichia coli.				
ACCESSION	BD268493.1	GI:33078261			
VERSION	BD268493.1	GI:33078261			
KEYWORDS	JP 2002535007-A/165.				
SOURCE	JP 2002535007-A/165.				
ORGANISM	unclassified.				
REFERENCE	1 (bases 1 to 1434)				
AUTHORS	Zyskind,J., Ohlsen,K.L., Trawick,J., Forsyth,A.R., Froelich,J.M., Carr,G.V., Yamamoto,R.T. and Xu,H.H.				
TITLE	Gene identified as being necessary in growth of Escherichia coli				
JOURNAL	Patent: JP 2002535007-A 165 22-OCT-2002;				
COMMENT	BLITRA PHARMACEUTICALS INC				
OS	B. coli				
PN	JP 2002535007-A/165				
PD	22-OCT-2002				
PF	27-JAN-2000				
PR	27-JAN-1999				
PI	JUDITH ZYSKIND, KARI L. OHLSEN, JOHN TRAWICK, ALLYN R. FORSYTH, PI JAMIE M. FROELICH.				
PI	GRANT J. CARR, ROBERT T. YAMAMOTO, HOWARD H. XU				
PC	C12N15/09, A61K31/7088, A61K48/00, A61P31/04, C07K14/245, C07K16/12, PC C12N1/15,				
PC	C12N1/19, C12N1/21, C12N5/10, C12P21/02, C12Q1/02, C12Q1/68 PC				
PC	G01N33/15, G01N33/50,				
PC	G01N33/53, G01N33/566, C12N15/00, C12N5/00				
CC	Gene identified as being necessary in growth of Escherichia				
CC	coll				
CC	Key				
CC	source				
CC	location/Qualifiers				
CC	1. 1434				
CC	/organism="B. coli".				
CC	location/Qualifiers				
CC	1. 1434				
CC	/organism="unclassified"				
CC	/mol_type="genomic DNA"				
CC	/db_xref="taxon:32644"				
FEATURES	source				
FEATURES	ORIGIN				



Oy 241 GCGGTAGTAAATCTGCGCGGAGTCAAGCTCAAAATGCGAATTGCTTCTGTAACGAC 300  
 Db 241 GCGGTAGTAAATCTGCGCGGAGTCAAGCTCAAAATGCGAATTGCTTCTGTAACGAC 300  
 Oy 301 CATCCGACATTAACCAATTACGGGTACTTTCCGCAACCAAGCTGATCCGTGGAT 360  
 Db 301 CATCCGACATTAACCAATTACGGGTACTTTCCGCAACCAAGCTGATCCGTGGAT 360  
 Oy 361 TTGGAAGAGTTTCAAGGTGTGATCGCAAGCCGCTGCAAGCCGATTAATCAGCG 420  
 Db 361 TTGGAAGAGTTTCAAGGTGTGATCGCAAGCCGCTGCAAGCCGATTAATCAGCG 420  
 Oy 421 CTGAGTTGATTTGCGCGCTGCTGTTCTGATACGCAAGGATGCGCTGGCAAGGTA 480  
 Db 421 CTGAGTTGATTTGCGCGCTGCTGTTCTGATACGCAAGGATGCGCTGGCAAGGTA 480  
 Oy 481 CAGCAGATGATCAACTGCGCGCTGTAAGCGGCTGTTCCGCTGCTGATTTCAAAAAGT 540  
 Db 481 CAGCAGATGATCAACTGCGCGCTGTAAGCGGCTGTTCCGCTGCTGATTTCAAAAAGT 540  
 Oy 541 ACCGATTTTGAAGCGCTACCGCGCGCTACGCTGTTAACGCGAATCTCTCGAATTTGAA 600  
 Db 541 ACCGATTTTGAAGCGCTACCGCGCGCTACGCTGTTAACGCGAATCTCTCGAATTTGAA 600  
 Oy 601 GGTGTGTGCTTAATGTAAGACGGAAGAAAGATTGTGAGCGCGGCAATGAACTGATT 660  
 Db 601 GGTGTGTGCTTAATGTAAGACGGAAGAAAGATTGTGAGCGCGGCAATGAACTGATT 660  
 Oy 661 GCCGATTAACGAATCTCGGCTCTGTAGTGAACCGCTTCCGAACAGGATATGCTGCTG 720  
 Db 661 GCCGATTAACGAATCTCGGCTCTGTAGTGAACCGCTTCCGAACAGGATATGCTGCTG 720  
 Oy 721 CAACCGGATTAAGCGCGCTGCTGATATGCAACCAAGCGCAAGATGTATGACGTTACC 780  
 Db 721 CAACCGGATTAAGCGCGCTGCTGATATGCAACCAAGCGCAAGATGTATGACGTTACC 780  
 Oy 781 GGTGTGTGCTTAATGTAAGACGGAAGAAAGATTGTGAGCGCGGCAATGAACTGATT 840  
 Db 781 GGTGTGTGCTTAATGTAAGACGGAAGAAAGATTGTGAGCGCGGCAATGAACTGATT 840  
 Oy 841 GAGAAACCTGCTCTTTGCAATGCGCGCTGCTGATGCTGCTGCAAACTGGGAAC 900  
 Db 841 GAGAAACCTGCTCTTTGCAATGCGCGCTGCTGATGCTGCTGCAAACTGGGAAC 900  
 Oy 901 TCCACGCTTTGCGCGATGACCTGGAATGCTGTGACGTGACGTGCAATACAGGCTTT 960  
 Db 901 TCCACGCTTTGCGCGATGACCTGGAATGCTGTGACGTGACGTGCAATACAGGCTTT 960  
 Oy 961 GCGCTGATGACCGGAAGGAACTGAAGCTGCGCTGATGCGGCAAGCGCTGAACCTGCTG 1020  
 Db 961 GCGCTGATGACCGGAAGGAACTGAAGCTGCGCTGATGCGGCAAGCGCTGAACCTGCTG 1020  
 Oy 1021 AAGTGTGATGACCAACGCTGCTTTGACATCTGCAACGCGCGGCAAGCTCTTATCTG 1080  
 Db 1021 AAGTGTGATGACCAACGCTGCTTTGACATCTGCAACGCGCGGCAAGCTCTTATCTG 1080  
 Oy 1081 GCAAAATGCTGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140  
 Db 1081 GCAAAATGCTGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140  
 Oy 1141 AAGTGTGATGACCAACGCTGCTTTGACATCTGCAACGCGCGGCAAGCTCTTATCTG 1200  
 Db 1141 AAGTGTGATGACCAACGCTGCTTTGACATCTGCAACGCGCGGCAAGCTCTTATCTG 1200  
 Oy 1201 GCGCAGATGAGAGGCTGCACTGCGGTAGTGTGTTGAGAGAGACAGCGCGCAGCGCTG 1260  
 Db 1201 GCGCAGATGAGAGGCTGCACTGCGGTAGTGTGTTGAGAGAGACAGCGCGCAGCGCTG 1260  
 Oy 1261 ATGCGCGGATCTTGGCAGATCTCTGTGTAAGAGCGCGGCACTATTAACCAAGAGATT 1320  
 Db 1261 ATGCGCGGATCTTGGCAGATCTCTGTGTAAGAGCGCGGCACTATTAACCAAGAGATT 1320

Oy 1321 GCCGAGTAAAGTCTGCGCGGAGTCAAGCTGCGGAAGTGTGCTCACTTGAAGAC 1380  
 Db 1321 GCCGAGTAAAGTCTGCGCGGAGTCAAGCTGCGGAAGTGTGCTCACTTGAAGAC 1380  
 Oy 1381 GGTGTGCTGACGACCAATCATCAAGAGATCAACAGGATTAATAAGGCTAA 1434  
 Db 1381 GGTGTGCTGACGACCAATCATCAAGAGATCAACAGGATTAATAAGGCTAA 1434  
 RESULT 3  
 LOCUS AR493607 1434 bp DNA linear PAT 15-MAY-2004  
 DEFINITION Sequence 165 from patent US 6720139.  
 ACCESSION AR493607  
 VERSION AR493607.1 GI:47265540  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (pages 1 to 1434)  
 AUTHORS Zykkind, J., Olsen, K.L., Trawick, J., Forsyth, R.A., Froelich, J.M.,  
 Carr, G.J., Yamamoto, R.T., and Xu, H.H.  
 TITL Genes identified as required for proliferation in *Bacterioides* coli  
 JOURNAL Patent: US 6720139-A 165 13-APR-2004;  
 Birta Pharmaceuticals, Inc.; San Diego, CA  
 FEATURES  
 source 1.1434  
 /organism="unknown"  
 /mol\_type="genomic DNA"  
 ORIGIN  
 Query Match 100.0%; Score 1434; DB 6; Length 1434;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 ATGAAAGTAAAGCTGCGGAGTGAACCTGCAAGTGAAGTGTGCTCACTTGAAGAC 60  
 Db 1 ATGAAAGTAAAGCTGCGGAGTGAACCTGCAAGTGAAGTGTGCTCACTTGAAGAC 60  
 Oy 61 CTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120  
 Db 61 CTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120  
 Oy 121 GTTAAAGTAAAGCTGCGGAGTGAACCTGCAAGTGAAGTGTGCTCACTTGAAGAC 180  
 Db 121 GTTAAAGTAAAGCTGCGGAGTGAACCTGCAAGTGAAGTGTGCTCACTTGAAGAC 180  
 Oy 181 GCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240  
 Db 181 GCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240  
 Oy 241 GCGCTGATTAATCTGCGCGGAGTCAAGCTGCAAGCTGCTGCTGCTGCTGCTG 300  
 Db 241 GCGCTGATTAATCTGCGCGGAGTCAAGCTGCAAGCTGCTGCTGCTGCTGCTG 300  
 Oy 301 CATCCGACATTAACCAATTACGGGTACTTTCCGCAACCAAGCTGATCCGTGGAT 360  
 Db 301 CATCCGACATTAACCAATTACGGGTACTTTCCGCAACCAAGCTGATCCGTGGAT 360  
 Oy 361 TTGGAAGAGTTTCAAGGTGTGATCGCAAGCCGCTGCAAGCCGATTAATCAGCG 420  
 Db 361 TTGGAAGAGTTTCAAGGTGTGATCGCAAGCCGCTGCAAGCCGATTAATCAGCG 420  
 Oy 421 CTGAGTTGATTTGCGCGCTGCTGTTCTGATACGCAAGGATGCGCTGGCAAGGTA 480  
 Db 421 CTGAGTTGATTTGCGCGCTGCTGTTCTGATACGCAAGGATGCGCTGGCAAGGTA 480  
 Oy 481 CAGCAGATGATCAACTGCGCGCTGTAAGCGGCTGTTCCGCTGCTGATTTCAAAAAGT 540  
 Db 481 CAGCAGATGATCAACTGCGCGCTGTAAGCGGCTGTTCCGCTGCTGATTTCAAAAAGT 540  
 Oy 541 ACCGATTTTGAAGCGCTACCGCGCGCTACGCTGTTAACGCGAATCTCTCGAATTTGAA 600

Db 541 ACCGATTTTGAGCCCTACCGCGCGCTACGCTGTAAACGCCGAATCTCTCGAATTTTGAA 600  
 Qy 601 GCTGTTCGGTAAATGATGACCGAAGAGAGATTTGTGAGCGCGCATGAAATCTGATT 660  
 Db 601 GCTGTTCGGTAAATGATGACCGAAGAGAGATTTGTGAGCGCGCATGAAATCTGATT 660  
 Qy 661 GCCGATTACGAATCTCGGCTCTGTAGTGAACCGCTTCCGACAGGGTATGTCGCTG 720  
 Db 661 GCCGATTACGAATCTCGGCTCTGTAGTGAACCGCTTCCGACAGGGTATGTCGCTG 720  
 Qy 721 CAACCGGGTAAAGCGCGCTGCTGATGCAACCAAGCGCGAGAGTGTATGAGCTTAAC 780  
 Db 721 CAACCGGGTAAAGCGCGCTGCTGATGCAACCAAGCGCGAGAGTGTATGAGCTTAAC 780  
 Qy 781 GGTGCGGGCGACAGGGTATTTGGCGTCTGCGCGCAACCGCTGAGAGGGGTAATTCGCTG 840  
 Db 781 GGTGCGGGCGACAGGGTATTTGGCGTCTGCGCGCAACCGCTGAGAGGGGTAATTCGCTG 840  
 Qy 841 GAAGAGCGCTGCTTCTTTGCAATGCGCGCGCTGCGCTGCTGCGCAACCTGGGAAC 900  
 Db 841 GAAGAGCGCTGCTTCTTTGCAATGCGCGCGCTGCGCTGCTGCGCAACCTGGGAAC 900  
 Qy 901 TCCACGGTTTCCGCGATGAGCTGAAATGCTGTAGTGAAGTGTGATGACGCTTT 960  
 Db 901 TCCACGGTTTCCGCGATGAGCTGAAATGCTGTAGTGAAGTGTGATGACGCTTT 960  
 Qy 961 GCGCGTATGACCGAAGAGAACTGAAAGCTGCGCGTGAAGCGCGCGTAAACGTTGAA 1020  
 Db 961 GCGCGTATGACCGAAGAGAACTGAAAGCTGCGCGTGAAGCGCGCGTAAACGTTGAA 1020  
 Qy 1021 AAAGTGTATGACCAACGCTGTCTTGAATCTGTGACGCGCGCGCACTCTTATCTG 1080  
 Db 1021 AAAGTGTATGACCAACGCTGTCTTGAATCTGTGACGCGCGCGCACTCTTATCTG 1080  
 Qy 1081 GCAAAATGCGCGAAGCTGAGTGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140  
 Db 1081 GCAAAATGCGCGAAGCTGAGTGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140  
 Qy 1141 AAACGGCTGAAAGGGGATTTCCCGCGCGGTAAACCACTGCAACAGCGTATGTTGCTG 1200  
 Db 1141 AAACGGCTGAAAGGGGATTTCCCGCGCGGTAAACCACTGCAACAGCGTATGTTGCTG 1200  
 Qy 1201 GCGCACTGAAAGCGTGTGATGCTGTGTTGAAGAGACACGCGCGACGCTTG 1260  
 Db 1201 GCGCACTGAAAGCGTGTGATGCTGTGTTGAAGAGACACGCGCGACGCTTG 1260  
 Qy 1261 ATCCCGGGATCTTCCGATCTGTGTGAAGCGCGCACTATTAACCAAGAGATT 1320  
 Db 1261 ATCCCGGGATCTTCCGATCTGTGTGAAGCGCGCACTATTAACCAAGAGATT 1320  
 Qy 1321 GCCGGAGTAAAGAGTCTGGGCGCAACGTTGTTGCTCACTTTGAAGAC 1380  
 Db 1321 GCCGGAGTAAAGAGTCTGGGCGCAACGTTGTTGCTCACTTTGAAGAC 1380  
 Qy 1381 GGTGCTGACGACCAATCATCAAGAGATCCAAAGATTAAGGCTAA 1434  
 Db 1381 GGTGCTGACGACCAATCATCAAGAGATCCAAAGATTAAGGCTAA 1434

RESULT 4  
 AX189075 1434 bp DNA linear PAT 08-AUG-2001  
 LOCUS Sequence 276 from Patent WO0148209.  
 DEFINITION AX189075  
 ACCESSION AX189075  
 VERSION AX189075.1 GI:15142798  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Escherichia coli  
 Escherichia coli  
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 Enterobacteriaceae; Escherichia.  
 REFERENCE  
 1 Forsyth, R.A., Ohlsen, K.L. and Zyakind, J.W.  
 Genes identified as required for proliferation of E. coli

JOURNAL Patent: WO 0148209-A 276 05-JUL-2001;  
 FEATURES Biftra Pharmaceuticals, Inc. (US)  
 source Location/Qualifiers  
 1..1434  
 /organism="Escherichia coli"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:562"  
 1..1434  
 CDS  
 /note="unnamed protein product"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAC50756.1"  
 /db\_xref="GI:15142798"  
 /translation="MKVTLPEPERAGVWVGVMDLDRVYGPTRISPEAPVVKVN  
 TIERPGLAANVAMVNIASLGAANALVGLGIDDAALSKSLADVVKCPVSPHP  
 TTKLAVSRNQILRLDPBEGREGDPPQPIHRIINQALSSIGALVLSQKAGLAVS  
 OQMTQLARKAPVILDPKGTDPBRRGATLTLPNISEBAVYKCKTEIEYRGSK  
 LADYELSLVTRSEQMSLIDPGAPLHPFLOQDSVIDVYAGDVIQVLAATLAA  
 GNSLEACFPANNAAGVVGKLTSTVSPLEENAVRGADVTGVTBSELALVAA  
 ARKRGKVVWTVNGVPDILHAGVSYLANAKLDRLIVAVNSDASTKRLGDSRPVNP  
 LEORMIVLGALEAVDVMVSPEDTPQLIAGLIPDLIVKGDYKPEBIAGSKWMANG  
 GEVLVINFEDGSTTNIKKIQDKKG"

ORIGIN  
 Query Match 100.0%; Score 1434; DB 6; Length 1434;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGAAAGTAAACGCTGCCAAGGTTTGAACGTGACGAGAGTATGTTGGTGTGATGATG 60  
 Db 1 ATGAAAGTAAACGCTGCCAAGGTTTGAACGTGACGAGAGTATGTTGGTGTGATGATG 60  
 Qy 61 CTGGAATCGTTACTGTATGACGCGCCCAACGATGATCTGCGCGGAAGCGCGTCCGCTG 120  
 Db 61 CTGGAATCGTTACTGTATGACGCGCCCAACGATGATCTGCGCGGAAGCGCGTCCGCTG 120  
 Qy 121 GTTAAAGTAAATCAATGAAAGCGTCCGCGCGCGCTTAACTGTCGATGATATTC 180  
 Db 121 GTTAAAGTAAATCAATGAAAGCGTCCGCGCGCGCTTAACTGTCGATGATATTC 180  
 Qy 181 GCTTCTCTCGGTCTAATGACCGCTGTCGCTGTTGACGCGGCAATGACATGACGCGCG 240  
 Db 181 GCTTCTCTCGGTCTAATGACCGCTGTCGCTGTTGACGCGGCAATGACATGACGCGCG 240  
 Qy 241 GCGCTGATTAATCTCTGCGCGACGTCACAGTCAAAATGCGACTTCTGTATCCGACG 300  
 Db 241 GCGCTGATTAATCTCTGCGCGACGTCACAGTCAAAATGCGACTTCTGTATCCGACG 300  
 Qy 301 CATCCGACATTAACCAATTAACGGGTATCTTCCCGCAACCAAGCTGATCCGTCTGAT 360  
 Db 301 CATCCGACATTAACCAATTAACGGGTATCTTCCCGCAACCAAGCTGATCCGTCTGAT 360  
 Qy 361 TTTGAAGAGGTTTCCAAAGGTGTTGATCCGCAACCGCTGACAGCGGATTAATCAGGCG 420  
 Db 361 TTTGAAGAGGTTTCCAAAGGTGTTGATCCGCAACCGCTGACAGCGGATTAATCAGGCG 420  
 Qy 421 CTGAGTTCGATTTGAGCGGCTGTGCTTCTGATACGCAAGGTGCGTCAAGGCTA 480  
 Db 421 CTGAGTTCGATTTGAGCGGCTGTGCTTCTGATACGCAAGGTGCGTCAAGGCTA 480  
 Qy 481 CAGCAATGATCCAACTGCGCGGTAAAGCGGGTGTTCGGTCTGATTAATCCAAAGGT 540  
 Db 481 CAGCAATGATCCAACTGCGCGGTAAAGCGGGTGTTCGGTCTGATTAATCCAAAGGT 540  
 Qy 541 ACCGATTTTGAAGGCTACCGCGCGCTACGCTGTAAACGCGAATCTCTCGAATTTGAA 600  
 Db 541 ACCGATTTTGAAGGCTACCGCGCGCTACGCTGTAAACGCGAATCTCTCGAATTTGAA 600  
 Qy 601 GCTGTTCGGTAAATGATGACCGAAGAGAGATTTGAGCGCGCATGAAATCTGATT 660  
 Db 601 GCTGTTCGGTAAATGATGACCGAAGAGAGATTTGAGCGCGCATGAAATCTGATT 660  
 Qy 661 GCCGATTACGAATCTCGGCTCTGTAGTGAACCGCTTCCGACAGGGTATGTCGCTG 720

Db 661 GCCATTACGAACTCTGGCTCTCTTACTGACCCGTTCCGACAGGGATATCCCTGCTG 720  
 Qy 721 CAACCGGGTAAAGCCGCTGTCATATGCAACCCAGCGCAGAAAGTATGACGTTACC 780  
 Db 721 CAACCGGGTAAAGCCGCTGTCATATGCAACCCAGCGCAGAAAGTATGACGTTACC 780  
 Qy 781 GGTGCGGGCCGACAGGTGATTTGGGCTCTTGGCGGCAACGCTGCGAGCGGGTAATTCGCTG 840  
 Db 781 GGTGCGGGCCGACAGGTGATTTGGGCTCTTGGCGGCAACGCTGCGAGCGGGTAATTCGCTG 840  
 Qy 841 GAAGAACCTCTCTCTTTGCAATGCGGCGGCTGGCGTGTGTCTGCGCAATGGGAACC 900  
 Db 841 GAAGAACCTCTCTCTTTGCAATGCGGCGGCTGGCGTGTGTCTGCGCAATGGGAACC 900  
 Qy 901 TCACCGGTTTCCCGATCGAGCTGGAATAATCTGTACGTGACGTGACATACAGGCTTT 960  
 Db 901 TCACCGGTTTCCCGATCGAGCTGGAATAATCTGTACGTGACGTGACATACAGGCTTT 960  
 Qy 961 GCGCTGATGACCGGAAGGAATCTGAAGTGGCCGTAGCGGCAAGCGCGTAAACGTGTGAA 1020  
 Db 961 GCGCTGATGACCGGAAGGAATCTGAAGTGGCCGTAGCGGCAAGCGCGTAAACGTGTGAA 1020  
 Qy 1021 AAAGTGTGATGACCAACGCTGTCTTTGACATCTGCAAGCGCGGCAAGCTCTTATCTG 1080  
 Db 1021 AAAGTGTGATGACCAACGCTGTCTTTGACATCTGCAAGCGCGGCAAGCTCTTATCTG 1080  
 Qy 1081 GCAAAATGCCCGCAAGCTGGGTGACCGCTTGATTTGGCCGTCAACAGGATGCTCCACC 1140  
 Db 1081 GCAAAATGCCCGCAAGCTGGGTGACCGCTTGATTTGGCCGTCAACAGGATGCTCCACC 1140  
 Qy 1141 AAAGGCTGAAAGGGGATTTCCCGCCCGGTAAACCCATCTCAACAGCGTATATGTGTG 1200  
 Db 1141 AAAGGCTGAAAGGGGATTTCCCGCCCGGTAAACCCATCTCAACAGCGTATATGTGTG 1200  
 Qy 1201 GCGGCACTGAAAGCGGTGCACTGGGTAGTGTCTTTGAAAGAGACACGCGCGCAGCGCTTG 1260  
 Db 1201 GCGGCACTGAAAGCGGTGCACTGGGTAGTGTCTTTGAAAGAGACACGCGCGCAGCGCTTG 1260  
 Qy 1261 ATGCCCGGAGCTTTGCCAGATCTCTGGTGAAGAGCGCGGCACTAATAACGAGAGATT 1320  
 Db 1261 ATGCCCGGAGCTTTGCCAGATCTCTGGTGAAGAGCGCGGCACTAATAACGAGAGATT 1320  
 Qy 1321 GCGCGGAGTAAAGAGTCTGGGCAACGCTGGCCGAAGTGTGTCTCAACTTTGAAGAC 1380  
 Db 1321 GCGCGGAGTAAAGAGTCTGGGCAACGCTGGCCGAAGTGTGTCTCAACTTTGAAGAC 1380  
 Qy 1381 GGTGTCTGACGACCAATCATCATCAAGAGATCCACAGATTAATAAGGCTTA 1434  
 Db 1381 GGTGTCTGACGACCAATCATCATCAAGAGATCCACAGATTAATAAGGCTTA 1434  
 RESULT 5  
 AX363670 1434 bp DNA linear PAT 15-FEB-2002  
 LOCUS AX363670  
 DEFINITION Sequence 165 from Patent EP1178052.  
 ACCESSION AX363670.1 GI:18695784  
 VERSION  
 KEYWORDS  
 SOURCE Escherichia coli  
 ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 Enterobacteriaceae; Escherichia.  
 REFERENCE 1  
 AUTHORS  
 TITLE Genes identified as required for proliferation in Escherichia coli  
 JOURNAL Patent: EP 1178052-A 165 06-FEB-2002;  
 FEATURES  
 SOURCE  
 1. 1434  
 /organism="Escherichia coli"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:562"

ORIGIN  
 Query Match 100.0%; Score 1434; DB 6; Length 1434;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1434; Conservative 0; Mismatches 0; Indels 0; Gaps 0  
 Qy 1 ATGAAAGTAAAGCGTCCAGGATTTGAACGTCGAGAGTGAATGGTGTGATGATGATG 60  
 Db 1 ATGAAAGTAAAGCGTCCAGGATTTGAACGTCGAGAGTGAATGGTGTGATGATGATG 60  
 Qy 61 CTGATCGTTACTGTGTAAGCGCCCAACGATCGTATCTGCGGGAAGCGCGGTCCGCTG 120  
 Db 61 CTGATCGTTACTGTGTAAGCGCCCAACGATCGTATCTGCGGGAAGCGCGGTCCGCTG 120  
 Qy 61 CTGATCGTTACTGTGTAAGCGCCCAACGATCGTATCTGCGGGAAGCGCGGTCCGCTG 120  
 Db 61 CTGATCGTTACTGTGTAAGCGCCCAACGATCGTATCTGCGGGAAGCGCGGTCCGCTG 120  
 Qy 121 GTTAAAGTAAATACATCGAAGAAAGTCCGCGGCGCGGCTTAACGTGGAGATGATATC 180  
 Db 121 GTTAAAGTAAATACATCGAAGAAAGTCCGCGGCGCGGCTTAACGTGGAGATGATATC 180  
 Qy 121 GTTAAAGTAAATACATCGAAGAAAGTCCGCGGCGCGGCTTAACGTGGAGATGATATC 180  
 Db 121 GTTAAAGTAAATACATCGAAGAAAGTCCGCGGCGCGGCTTAACGTGGAGATGATATC 180  
 Qy 181 GCTTCTCTGCGGTCTAAATGACGCTGGTGGGTGACGCGGCAATGACATGACGCGC 240  
 Db 181 GCTTCTCTGCGGTCTAAATGACGCTGGTGGGTGACGCGGCAATGACATGACGCGC 240  
 Qy 181 GCTTCTCTGCGGTCTAAATGACGCTGGTGGGTGACGCGGCAATGACATGACGCGC 240  
 Db 181 GCTTCTCTGCGGTCTAAATGACGCTGGTGGGTGACGCGGCAATGACATGACGCGC 240  
 Qy 241 GCGCTGAGTAAATCTCTGCGCGACGTCACAGTCAATGCGACTTCTGTACCAACG 300  
 Db 241 GCGCTGAGTAAATCTCTGCGCGACGTCACAGTCAATGCGACTTCTGTACCAACG 300  
 Qy 241 GCGCTGAGTAAATCTCTGCGCGACGTCACAGTCAATGCGACTTCTGTACCAACG 300  
 Db 241 GCGCTGAGTAAATCTCTGCGCGACGTCACAGTCAATGCGACTTCTGTACCAACG 300  
 Qy 301 CATCCGACATTAACCAATTAACGAGTATCTTCCGCAACCAACAGCTGATCCGTGAT 360  
 Db 301 CATCCGACATTAACCAATTAACGAGTATCTTCCGCAACCAACAGCTGATCCGTGAT 360  
 Qy 301 CATCCGACATTAACCAATTAACGAGTATCTTCCGCAACCAACAGCTGATCCGTGAT 360  
 Db 301 CATCCGACATTAACCAATTAACGAGTATCTTCCGCAACCAACAGCTGATCCGTGAT 360  
 Qy 361 TTTGAAGAGTTTCAAGAGTGTGATCCGACGCGCTGACAGAGCGGATTAATCAAGG 420  
 Db 361 TTTGAAGAGTTTCAAGAGTGTGATCCGACGCGCTGACAGAGCGGATTAATCAAGG 420  
 Qy 361 TTTGAAGAGTTTCAAGAGTGTGATCCGACGCGCTGACAGAGCGGATTAATCAAGG 420  
 Db 361 TTTGAAGAGTTTCAAGAGTGTGATCCGACGCGCTGACAGAGCGGATTAATCAAGG 420  
 Qy 421 CTGAGTTCGATTTGCGCGCTGCTGCTTTCTGACTACGCAAGGTGCGTGGCAAGCTGA 480  
 Db 421 CTGAGTTCGATTTGCGCGCTGCTGCTTTCTGACTACGCAAGGTGCGTGGCAAGCTGA 480  
 Qy 421 CTGAGTTCGATTTGCGCGCTGCTGCTTTCTGACTACGCAAGGTGCGTGGCAAGCTGA 480  
 Db 421 CTGAGTTCGATTTGCGCGCTGCTGCTTTCTGACTACGCAAGGTGCGTGGCAAGCTGA 480  
 Qy 481 CAGCAGATGATCCAACTGGCGCGTAAAGCGGCTGCTCCGCTGATGATTCGAAAGGT 540  
 Db 481 CAGCAGATGATCCAACTGGCGCGTAAAGCGGCTGCTCCGCTGATGATTCGAAAGGT 540  
 Qy 481 CAGCAGATGATCCAACTGGCGCGTAAAGCGGCTGCTCCGCTGATGATTCGAAAGGT 540  
 Db 481 CAGCAGATGATCCAACTGGCGCGTAAAGCGGCTGCTCCGCTGATGATTCGAAAGGT 540  
 Qy 541 ACCGATTTGAGGCTACGCGCGGCGCTAGCTGTAAAGCGCAATCTCGGAAATTTGA 600  
 Db 541 ACCGATTTGAGGCTACGCGCGGCGCTAGCTGTAAAGCGCAATCTCGGAAATTTGA 600  
 Qy 541 ACCGATTTGAGGCTACGCGCGGCGCTAGCTGTAAAGCGCAATCTCGGAAATTTGA 600  
 Db 541 ACCGATTTGAGGCTACGCGCGGCGCTAGCTGTAAAGCGCAATCTCGGAAATTTGA 600  
 Qy 601 GCTGTGTCCGTAATGTAAGCAAGAGAGATTTGTAAGCGCGCATGAATGAT 660  
 Db 601 GCTGTGTCCGTAATGTAAGCAAGAGAGATTTGTAAGCGCGCATGAATGAT 660  
 Qy 601 GCTGTGTCCGTAATGTAAGCAAGAGAGATTTGTAAGCGCGCATGAATGAT 660  
 Db 601 GCTGTGTCCGTAATGTAAGCAAGAGAGATTTGTAAGCGCGCATGAATGAT 660  
 Qy 661 GCGGATTAAGCACTCTCGGCTCTGTTAGTGAACCGCTTCCGAAAGGATATGCTGCTG 720  
 Db 661 GCGGATTAAGCACTCTCGGCTCTGTTAGTGAACCGCTTCCGAAAGGATATGCTGCTG 720  
 Qy 661 GCGGATTAAGCACTCTCGGCTCTGTTAGTGAACCGCTTCCGAAAGGATATGCTGCTG 720  
 Db 661 GCGGATTAAGCACTCTCGGCTCTGTTAGTGAACCGCTTCCGAAAGGATATGCTGCTG 720  
 Qy 721 CAACCGGGTAAAGCGCGCTGCAATGCAACCCAGCGCAGAAAGTATGACGTTACC 780  
 Db 721 CAACCGGGTAAAGCGCGCTGCAATGCAACCCAGCGCAGAAAGTATGACGTTACC 780  
 Qy 721 CAACCGGGTAAAGCGCGCTGCAATGCAACCCAGCGCAGAAAGTATGACGTTACC 780  
 Db 721 CAACCGGGTAAAGCGCGCTGCAATGCAACCCAGCGCAGAAAGTATGACGTTACC 780  
 Qy 781 GGTGCGGGCCGACAGGTGATTTGGGCTCTTGGCGGCAACGCTGCGAGCGGGTAATTCGCTG 840  
 Db 781 GGTGCGGGCCGACAGGTGATTTGGGCTCTTGGCGGCAACGCTGCGAGCGGGTAATTCGCTG 840  
 Qy 781 GGTGCGGGCCGACAGGTGATTTGGGCTCTTGGCGGCAACGCTGCGAGCGGGTAATTCGCTG 840  
 Db 781 GGTGCGGGCCGACAGGTGATTTGGGCTCTTGGCGGCAACGCTGCGAGCGGGTAATTCGCTG 840  
 Qy 841 GAAGAACCTCTCTCTTTGCAATGCGGCGGCTGGCGTGTGTCTGCGCAATGGGAACC 900  
 Db 841 GAAGAACCTCTCTCTTTGCAATGCGGCGGCTGGCGTGTGTGTCTGCGCAATGGGAACC 900  
 Qy 841 GAAGAACCTCTCTCTTTGCAATGCGGCGGCTGGCGTGTGTGTCTGCGCAATGGGAACC 900  
 Db 841 GAAGAACCTCTCTCTTTGCAATGCGGCGGCTGGCGTGTGTGTCTGCGCAATGGGAACC 900  
 Qy 901 TCACCGGTTTCCCGATCGAGCTGGAATAATCTGTACGTGACGTGACATACAGGCTTT 960  
 Db 901 TCACCGGTTTCCCGATCGAGCTGGAATAATCTGTACGTGACGTGACATACAGGCTTT 960  
 Qy 901 TCACCGGTTTCCCGATCGAGCTGGAATAATCTGTACGTGACGTGACATACAGGCTTT 960  
 Db 901 TCACCGGTTTCCCGATCGAGCTGGAATAATCTGTACGTGACGTGACATACAGGCTTT 960  
 Qy 961 GCGCTGATGACCGGAAGGAATCTGAAGTGGCCGTAGCGGCAAGCGCGTAAACGTGTGAA 1020  
 Db 961 GCGCTGATGACCGGAAGGAATCTGAAGTGGCCGTAGCGGCAAGCGCGTAAACGTGTGAA 1020

Oy		1021	AAAGGGGATGAGCAACGAGTCTTTGACATCCTGCAGCGCCGGGACGTCTTAATCG	1080
Dd		1021	AAAGTGATGATGCCAACGGATGTCCTTTGACATCTGCAGCGCCGGGACGTCTTAATCG	1080
Oy		1081	GCAAATGCGCGCACACTGGTGGTAGCCGCTTGATTGTTGCGGTCAACAGCATGCTCCACC	1140
Dd		1081	GCAAATGCGCGCACAGCTGGTAGCCGCTTGATTGTTGCGGTCAACAGCATGCTCCACC	1140
Oy		1141	AAAAGGCTGAAAAGGGGATTTCCCOCGGGTAAACCACTCGAACAAGCTTATGTTGCTG	1200
Dd		1141	AAAAGGCTGAAAAGGGGATTTCCCOCGGGTAAACCACTCGAACAAGCTTATGTTGCTG	1200
Oy		1201	GGCGCACTGGAAGCGGTGCACTGGGTAGTGTGCTTGAAGAAGACACCCCGCAGCGCTTG	1260
Dd		1201	GGCGCACTGGAAGCGGTGCACTGGGTAGTGTGCTTGAAGAAGACACCCCGCAGCGCTTG	1260
Oy		1261	ATCGCCGGGATCTTGCCAGATCTGCTGTGTGAAGGCGGCGACTATTAACAGAAAGATT	1320
Dd		1261	ATCGCCGGGATCTTGCCAGATCTGCTGTGTGAAGGCGGCGACTATTAACAGAAAGATT	1320
Oy		1321	GCCGGAGATAAAGAAAGTCTGGGCGCAACGATGCGAAAGTGTGGTCAACTTGAAGAC	1380
Dd		1321	GCCGGAGATAAAGAAAGTCTGGGCGCAACGATGCGAAAGTGTGGTCAACTTGAAGAC	1380
Oy		1381	GATTGCTCGAGCAGCACCATCATCAAGAAAGATCCACAGATTAATAAAGGCTAA	1434
Dd		1381	GATTGCTCGAGCAGCACCATCATCAAGAAAGATCCACAGATTAATAAAGGCTAA	1434
RESULT 6				
ECU28379/c				
LOCUS		25638 bp	DNA	linear BCT 27-FEB-1996
DEFINITION		Escherichia coli K-12 genome; approximately 68 minutes.		
VERSION		U28379		
ACCESSION		U28379.1		GI:1203797
KEYWORDS				
SOURCE				
ORGANISM		Escherichia coli		
		Escherichia coli		
		Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
		Enterobacteriaceae; Escherichia.		
		1 (bases 1 to 25638)		
REFERENCE				
AUTHORS		Plunkett,G.		
TITLE		Direct Submision		
JOURNAL		Submitted (01-JUN-1995) Guy Plunkett, Laboratory of Genetics,		
		University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA		
		On Feb 27, 1996 this sequence version replaced gi:882576		
COMMENT		This sequence was determined as part of the E. coli Genome Project		
		(Frederick R. Blattner, director) at the University of		
		Wisconsin-Madison. Supported by award H000301 from the NIH Human		
		Genome Project. The entire sequence was independently determined		
		from E. coli MG1655; overlaps with other sequence determinations		
		are annotated. The end of this entry overlaps the start of the		
		entry ECOW67 (U18957) by 1596 bp. This entry should be considered		
		somewhat provisional; it will be updated and merged with others at		
		a later date.		
FEATURES				
source		Location/Qualifiers		
		1..25638		
		/organism="Escherichia coli"		
		/mol_type="genomic DNA"		
		/strain="K-12"		
		/sub_strain="MG1655"		
		/db_xref="taxon:562"		
		/map="approximately 68 minutes"		
		/note="This sequence is from lambda clone		
		DD689(EC27-120); MyDanns vector was used for subcloning		
		6002..6622		
		/note="alternate name ygiM; ORF_o206"		
		/codon_start=1		
		/transl_table=1		
		/protein_id="AAA89135.1"		
		/db_xref="gi:1203798"		
		/translation="WPKRLIGLTLLASATAVSAESTRYVSDLNTWVRSGEDH		
CDS				

[illegible]

misc\_feature 10928..15584  
/note="similar to GenBank Accession Number M16194  
(ECORPSU)"  
10978..15584  
/note="similar to GenBank Accession Number L14781  
(ECOTATDAB)"  
gene 11258..12169  
/gene="tcdA"  
11258..12169  
/gene="tcdA"  
CDS 11258..12169  
/gene="tcdA"  
/BC number="4.2.1.32"  
/codon\_start=1  
/transl\_table=1  
/product="tartarate dehydratase"  
/protein\_id="AA089141.1"  
/db\_xref="GI:882584"  
/translacion="MMSBNKQAVNKLTEIVANFTAMISTREDDVDLKKQKAE  
TSGMKIYHTMENDMOKAIDLRNPAQDGTGIMFVKVSGRPLLEGLQSLKQAVE  
EATYKAPLRNNAVEIPDEVNTKNTSGVPMWMDIIPNDDBAIEIVYAGGGCTLP  
RSYVMPSEGEVGVKVPVNPISITLNAACPVLYGVGIVSTVETAIVSRKALIRI  
GSRHNPKAAELERLEEGNRLGICPGILGNSVNGVHISBAARHPTTGVAVSTG  
CMARRGTLLVHDLPENLSHTSAL"  
12166..12771  
/gene="tcdB"  
12166..12771  
/gene="tcdB"  
CDS 12166..12771  
/gene="tcdB"  
/BC number="4.2.1.32"  
/codon\_start=1  
/transl\_table=1  
/product="tartarate dehydratase"  
/protein\_id="AA089142.1"  
/db\_xref="GI:882585"  
/translacion="MKKILTPPIKAEDLDIRVGVYVLTGLVTCDDVCHRLIEK  
RPIPYDLNGKAIFPAGPIVRKNGDKMENVGPTTSRMSPFERFLEQTVLXVEK  
GPIVSIIDTGNMLIAENKCLFARRDPIVEICEHVIYIK"  
12819..14282  
/note="alternate name y5jB; ORF\_0487"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="AA09143.1"  
/db\_xref="GI:882586"  
/translacion="MKPSTMMRYLAPLAVIAITALLPVPAGLHNTMLYPAFTGVY  
VGLIRPVPNAVVMGISTITILSPBLNPGPKTKAKSLMVASPSNV  
IMLIFAPMFETGETGELGRILNLIKMGHRTLPGLYVMSBILLPVPSNSA  
RGAGIYPIIRNLPPLVQSPNDSSRSISYIMMGIADVTSALFLVAPNELL  
IGLMSASHATLSGDMFLQMLPLSILLVLPVLAVALVPPVLSKSDQVPRABTEL  
OAMGPLCSREKRMGLMVALVLMIFGSDVLDAMVGVSVLALMLLRLISMPDIYSN  
KAMNVFPMILALITLALGNNGPIFISPKILAGSLSGVSPMMVALIVPDIILRY  
PFSATVYTSALAPMMLAALAMPRIPLPFCMLMGAIGLSILPTATGSPPIYTG  
SGYLPADVRLGIFGLIFLVLLVITGLMMPVLL"  
complement (14325..15338)  
/note="ORF\_0337"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="AA09144.1"  
/db\_xref="GI:882587"  
/translacion="MVALIGIESCTETGIAIYDEKGLANLYSQVKLADYGGVVP  
ELASRDVARTVPLIQALKESTGLADIADVATAPGVLGALVATAGRSALFAM  
DVPAIPVHNEGHLLAPMLBNDPSPFVLLVSGGTOLISVTGICQYLLGSIDP  
AABGAPKTAKGLGLDYRGGPILSKMAAGTAGFVRPMTBRTGIDPSPGLKIPA  
ANTIRNGTDQTRADYARAFEDAVYDTLMIKCRALDQGFPRKLVMAAGVANSRTIR  
AKLAEMMKRRGEVYFARPECTDNGMIAIVAGNVRKAATADLVGVSVRPMLLEL  
PAA"  
misc\_feature 14862..15917  
/note="similar to GenBank Accession Number J01687  
(ECORPSRPO)"  
15576..15791  
/gene="rpbU"  
15576..15791  
/gene="rpbU"  
CDS 15576..15791  
/note="CG site No. 211"  
/codon\_start=1

Query Match 100.0%; Score 1434; DB 1; Length 25638;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
/transl\_table=1  
/product="30S ribosomal subunit protein S21"  
/protein\_id="AA089145.1"  
/db\_xref="GI:882588"  
/translacion="MPVIVRENEPPDALRRPKSCERAGVLAVERREBEKPTTE  
RRKAKASAVRPAKLAARENARTRLY"  
gene 15902..17647  
/gene="dnaG"  
15902..17647  
/gene="dnaG"  
CDS 15902..17647  
/function="DNA biosynthesis; primase"  
/note="CG Site No. 847; alternate gen name dnaP, parB"  
/codon\_start=1  
Query Match 100.0%; Score 1434; DB 1; Length 25638;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 ATGAAAGTAAACGCTGCAGAGTTTGAACGTGCAGAGTGAATGGTGGTGAATGATG 60  
1548 ATGAAAGTAAACGCTGCAGAGTTTGAACGTGCAGAGTGAATGGTGGTGAATGATG 1489  
61 CTGATCGTTACTGTGTAACGCGCCACAGTGTATCTCGCCGGAAGCGCGTCCCGTG 120  
1488 CTGATCGTTACTGTGTAACGCGCCACAGTGTATCTCGCCGGAAGCGCGTCCCGTG 1429  
121 GTTAAAGTAAACGCTGCAGAGTTTGAACGTGCAGAGTGAATGGTGGTGAATGATG 180  
1428 GTTAAAGTAAACGCTGCAGAGTTTGAACGTGCAGAGTGAATGGTGGTGAATGATG 1369  
181 GCTTCTCGGTGCTATGACAGCGCTGTGTGGGTTGACGGGCAATTGACATGACGCGCG 240  
1368 GCTTCTCGGTGCTATGACAGCGCTGTGTGGGTTGACGGGCAATTGACATGACGCGCG 1309  
241 GCGCTAGTAATCTGTGCGCAAGTCAAGTCAATGCGAATGCGATCTGTTCTGACGACG 300  
1308 GCGCTAGTAATCTGTGCGCAAGTCAAGTCAATGCGAATGCGATCTGTTCTGACGACG 1249  
301 CATCCGACCTTACCAATTAACGGGTACTTCCCGCAACCAACGCTGATCCGCTCGAT 360  
1248 CATCCGACCTTACCAATTAACGGGTACTTCCCGCAACCAACGCTGATCCGCTCGAT 1189  
361 TTGGAAGAGGTTTGAAGGTTTGAATCCGACCGCTGACAGAGCGAATTAATCAGCGG 420  
1188 TTGGAAGAGGTTTGAAGGTTTGAATCCGACCGCTGACAGAGCGAATTAATCAGCGG 1129  
421 CTGAGTTGCAATTTGCGCGCTGTGCTTCTGACTACGCAAGGTGCGTGGCAAGGTA 480  
1128 CTGAGTTGCAATTTGCGCGCTGTGCTTCTGACTACGCAAGGTGCGTGGCAAGGTA 1069  
481 CAGCAGATGATCCACTGCGCGGTAAAGCGGGTGTCCGCTGCTGATTAATCCAAAAGT 540  
1068 CAGCAGATGATCCACTGCGCGGTAAAGCGGGTGTCCGCTGCTGATTAATCCAAAAGT 1009  
541 ACCGATTTGAGGGCTACACGCGCGGTAAAGCGGGTGTAAAGCGGAAATCTCGAATTTGAA 600  
1008 ACCGATTTGAGGGCTACACGCGCGGTAAAGCGGGTGTAAAGCGGAAATCTCGAATTTGAA 949  
601 GCTGTGTCGTAATATGTAAGACGGAAGAGATTGTAAGCGCGCATGAATGAT 660  
948 GCTGTGTCGTAATATGTAAGACGGAAGAGATTGTAAGCGCGCATGAATGAT 989  
661 GCGGATTAAGATCTTCGCGTCTGTTAGTAAACCGGTTCCGAACAGGGTATGTCGCTG 720  
888 GCGGATTAAGATCTTCGCGTCTGTTAGTAAACCGGTTCCGAACAGGGTATGTCGCTG 829  
721 CAACCGGGTAAAGCGCGCTGATATGCAACCAAGCGCAGGAAGTATGATGATAC 780  
828 CAACCGGGTAAAGCGCGCTGATATGCAACCAAGCGCAGGAAGTATGATGATAC 769  
781 GGTGCGGGCGACACGCTGATTTGGCGTCTGCGCGCAACGCTGGCAGCGGGTAAATTCGCTG 840  
768 GGTGCGGGCGACACGCTGATTTGGCGTCTGCGCGCAACGCTGGCAGCGGGTAAATTCGCTG 709



QY 841 GAAGAGCTCTCTTTGCGCAATGCGCGCTGCGGTGATCGGCAAACTGGAAACC 900  
 DB 708 GAAGAAGCTCTCTTTGCGCAATGCGCGCTGCGGTGATCGGCAAACTGGAAACC 649  
 QY 901 TCACGCGTTTGGCGATCGAGCTGAAAATGCTGTACGTGACGTGCAATACAGCTTT 960  
 DB 648 TCACGCGTTTGGCGATCGAGCTGAAAATGCTGTACGTGACGTGCAATACAGCTTT 589  
 QY 961 GCGCGTATGACCGGAAGGAATGAAAGTGGCCGTAGCGGAGGCGGTAAACGTGGGAA 1020  
 DB 588 GCGCGTATGACCGGAAGGAATGAAAGTGGCCGTAGCGGAGGCGGTAAACGTGGGAA 529  
 QY 1021 AAGTGTGATGACCAACGAGTGTCTTTGACATCTCTGACGCGCGGACGCTCTTATCTG 1080  
 DB 528 AAGTGTGATGACCAACGAGTGTCTTTGACATCTCTGACGCGCGGACGCTCTTATCTG 469  
 QY 1081 GCAAAATGCCCGAAGCTGGGTGACCGCTGTATGTTGCGGTCAACAGCGATGCTCCACC 1140  
 DB 468 GCAAAATGCCCGAAGCTGGGTGACCGCTGTATGTTGCGGTCAACAGCGATGCTCCACC 409  
 QY 1141 AAGCGCTGAAGAGGAGTTCGCGCGCGGTAAACCACTCGAACAAGCGTATGATGCTG 1200  
 DB 408 AAGCGCTGAAGAGGAGTTCGCGCGCGGTAAACCACTCGAACAAGCGTATGATGCTG 349  
 QY 1201 GCGGCACTGAAGCGGTGACCTGGGTAGTGTCTTTGAAAGAGACAGCGCGACGCTTG 1260  
 DB 348 GCGGCACTGAAGCGGTGACCTGGGTAGTGTCTTTGAAAGAGACAGCGCGACGCTTG 289  
 QY 1261 ATGCGCGGAGTCTTGGCAGATCTGCTGTGTAAGAGCGCGGCACTATTAACCAAGAGATT 1320  
 DB 288 ATGCGCGGAGTCTTGGCAGATCTGCTGTGTAAGAGCGCGGCACTATTAACCAAGAGATT 229  
 QY 1321 GCGCGAGTAAAGAGTCTGGGCGCAACGCTGGCGAAGCTGTGGCTCAACTTTGAAGAC 1380  
 DB 228 GCGCGAGTAAAGAGTCTGGGCGCAACGCTGGCGAAGCTGTGGCTCAACTTTGAAGAC 169  
 QY 1381 GGTGTCTGACGACCAATCATCATCAAGAGATCCACAGATTAAGAGCTAA 1434  
 DB 168 GGTGTCTGACGACCAATCATCATCAAGAGATCCACAGATTAAGAGCTAA 115

RESULT 7  
 U00096 31/c  
 WPCOMMENT

Sequence split into 47 fragments LOCUS U00096 Accession U00096

Fragment Name	Begin	End
U00096_00	1	110000
U00096_01	100001	210000
U00096_02	200001	310000
U00096_03	300001	410000
U00096_04	400001	510000
U00096_05	500001	610000
U00096_06	600001	710000
U00096_07	700001	810000
U00096_08	800001	910000
U00096_09	900001	1010000
U00096_10	1000001	1110000
U00096_11	1100001	1210000
U00096_12	1200001	1310000
U00096_13	1300001	1410000
U00096_14	1400001	1510000
U00096_15	1500001	1610000
U00096_16	1600001	1710000
U00096_17	1700001	1810000
U00096_18	1800001	1910000
U00096_19	1900001	2010000
U00096_20	2000001	2110000
U00096_21	2100001	2210000
U00096_22	2200001	2310000
U00096_23	2300001	2410000
U00096_24	2400001	2510000
U00096_25	2500001	2610000

U00096\_26 2600001 2710000  
 U00096\_27 2700001 2810000  
 U00096\_28 2800001 2910000  
 U00096\_29 2900001 3010000  
 U00096\_30 3000001 3110000  
 U00096\_31 3100001 3210000  
 U00096\_32 3200001 3310000  
 U00096\_33 3300001 3410000  
 U00096\_34 3400001 3510000  
 U00096\_35 3500001 3610000  
 U00096\_36 3600001 3710000  
 U00096\_37 3700001 3810000  
 U00096\_38 3800001 3910000  
 U00096\_39 3900001 4010000  
 U00096\_40 4000001 4110000  
 U00096\_41 4100001 4210000  
 U00096\_42 4200001 4310000  
 U00096\_43 4300001 4410000  
 U00096\_44 4400001 4510000  
 U00096\_45 4500001 4610000  
 U00096\_46 4600001 4639675

Continuation (32 of 47) of U00096 from base 3100001 (U00096 Escherichia coli K-12 MG1655)

Query Match 100.0%; Score 1434; DB 1; Length 110000;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAAGTAAAGCTGCGCAGGTTTGAACGTGACGAGATGATGGTGTGATGTGATG 60  
 DB 94775 ATGAAGTAAAGCTGCGCAGGTTTGAACGTGACGAGATGATGGTGTGATGTGATG 94716  
 QY 61 CTGATCGTACTGTGATGAGCGCCCAACAGTCTATCTCGCCGGAAGCGCGTCCGCTG 120  
 DB 94715 CTGATCGTACTGTGATGAGCGCCCAACAGTCTATCTCGCCGGAAGCGCGTCCGCTG 94656  
 QY 121 GTTAAAGTAAATCCATCGAAGAAAGTCCGGGCGCGGCTTAACTGTGGATGATATC 180  
 DB 94655 GTTAAAGTAAATCCATCGAAGAAAGTCCGGGCGCGGCTTAACTGTGGATGATATC 94596  
 QY 181 GCTTCTCGGTGCTATGACGCGTGGTGGGTGACGGGCAATTGACATGACGCGCG 240  
 DB 94595 GCTTCTCGGTGCTATGACGCGTGGTGGGTGACGGGCAATTGACATGACGCGCG 94536  
 QY 241 GCGCTGAGTAAATCTCGGCGGACGTCAACGTCAAAATGCGACTTCTGTACCGACG 300  
 DB 94535 GCGCTGAGTAAATCTCGGCGGACGTCAACGTCAAAATGCGACTTCTGTGTACCGACG 94476  
 QY 301 CATCCGACATTACCAATTAACGGGTACTTCCCGCAACCAAGCTGATCCGTCTGGAT 360  
 DB 94475 CATCCGACATTACCAATTAACGGGTACTTCCCGCAACCAAGCTGATCCGTCTGGAT 94416  
 QY 361 TTTGAAGAAAGTTTGAAGGTTGATCCGACGCGTGCACGAGCGATTAATCAGGCG 420  
 DB 94415 TTTGAAGAAAGTTTGAAGGTTGATCCGACGCGTGCACGAGCGATTAATCAGGCG 94356  
 QY 421 CTGAGTTCGATTTGCGCGCTGTGCTTCTGACTACGCAACCAAGGTGCTGCAAGGCTA 480  
 DB 94355 CTGAGTTCGATTTGCGCGCTGTGCTTCTGACTACGCAACCAAGGTGCTGCAAGGCTA 94296  
 QY 481 CAGCAGATGATCCACCTGGCGGCTTAAAGCGGCTGTCCGCTGATTAATCCAAAAGGT 540  
 DB 94295 CAGCAGATGATCCACCTGGCGGCTTAAAGCGGCTGTCCGCTGATTAATCCAAAAGGT 94236  
 QY 541 ACCGATTTTGAAGGCTACCGCGGCGCTGACGCTTAAACGCGAATCTCGGAATTTGAA 600  
 DB 94235 ACCGATTTTGAAGGCTACCGCGGCGCTGACGCTTAAACGCGAATCTCGGAATTTGAA 94176  
 QY 601 GCTGTGTGCGTAAATGTAAGACCGAAGAAAGATTTGAGCGCGCATGAATCTGATTT 660  
 DB 94175 GCTGTGTGCGTAAATGTAAGACCGAAGAAAGATTTGAGCGCGCATGAATCTGATTT 94116  
 QY 661 GCCGATTAAGAACTTCGCGCTCTGTATGATCCCGTTCCGAACGAGGTATGTGCTGCTG 720



```

Db      94115 GCGGATTACGACTCTGGCTCTGTAGTACCCGTTCCGAAACGGGTATGCTGCTG 94056
Qy      721 CAACCGGGTAAAGCGCGCTGCAATATGCCAACCGGACGAGAGTATGACGTTACC 780
Db      94055 CAACCGGGTAAAGCGCGCTGCAATATGCCAACCGGACGAGAGTATGACGTTACC 93996
Qy      781 GGTGCGGGGCAACGCGGTATTTGGGCTCTGGCGGCAACGCTGGACGAGGGGTATTTGGCTG 840
Db      93995 GGTGCGGGGCAACGCGGTATTTGGGCTCTGGCGGCAACGCTGGACGAGGGGTATTTGGCTG 93936
Qy      841 GAAGAACCTCTCTCTTTGCAATGCGCGCTGCGCTGGTGTCTGGCAATCTGGGAACC 900
Db      93935 GAAGAACCTCTCTCTTTGCAATGCGCGCTGCGCTGGTGTCTGGCAATCTGGGAACC 93876
Qy      901 TCACGCGTTTCTGCGGATGCACTGCAAAATGCTGTACGTGACGTGCAATACAGGCTTT 960
Db      93875 TCACGCGTTTCTGCGGATGCACTGCAAAATGCTGTACGTGACGTGCAATACAGGCTTT 93816
Qy      961 GGGGTGATGACCGGAAGGAACCTGAAGCTGGCGGCTAGCGGACGCGTAAACGTGTGAA 1020
Db      93815 GGGGTGATGACCGGAAGGAACCTGAAGCTGGCGGCTAGCGGACGCGTAAACGTGTGAA 93756
Qy      1021 AAGGTGATGACCAACAGGTGTCTTTGACATCTCTGACGCGCGGCAAGTCTTTATCTG 1080
Db      93755 AAGGTGATGACCAACAGGTGTCTTTGACATCTCTGACGCGCGGCAAGTCTTTATCTG 93696
Qy      1081 GCAATATCCCGCAAGCTGGGTGACCGCTTGAATTTGTCCTTCAACAGCGATGCTTCAAC 1140
Db      93695 GCAATATCCCGCAAGCTGGGTGACCGCTTGAATTTGTCCTTCAACAGCGATGCTTCAAC 93636
Qy      1141 AAGCGGTGAAGGGGGATTCCCGCGCGGTAAACCACTCGAACGCGTATGATTGCTG 1200
Db      93635 AAGCGGTGAAGGGGGATTCCCGCGCGGTAAACCACTCGAACGCGTATGATTGCTG 93576
Qy      1201 GGGCACTGGAAGCGGTGCACTGGGTAGTGTCTTTGAAGAGACACGCGGACGCGCTTG 1260
Db      93575 GGGCACTGGAAGCGGTGCACTGGGTAGTGTCTTTGAAGAGACACGCGGACGCGCTTG 93516
Qy      1261 ATGCGCGGGATCTTGGCAAGATCTGCTGTGTGAAGGCGGCGCATTAACCAAGAGATT 1320
Db      93515 ATGCGCGGGATCTTGGCAAGATCTGCTGTGTGAAGGCGGCGCATTAACCAAGAGATT 93456
Qy      1321 GCGGGAGTAAAGAGTCTGGGCGCAAGGTGGGAGAGTGTGGTCTCAACTTTGAAGAC 1380
Db      93455 GCGGGAGTAAAGAGTCTGGGCGCAAGGTGGGAGAGTGTGGTCTCAACTTTGAAGAC 93396
Qy      1381 GGTGTCTGACGACCAACATCATCAAGAGATCCCAAGAGTAAAAAGGCTAA 1434
Db      93395 GGTGTCTGACGACCAACATCATCAAGAGATCCCAAGAGTAAAAAGGCTAA 93342

```

# RESULT 8 AE005174\_39/c

Sequence split into 56 fragments, LOCUS AE005174 Accession AE005174

Fragment Name	Begin	End
AE005174_00	1	110000
AE005174_01	100001	210000
AE005174_02	200001	310000
AE005174_03	300001	410000
AE005174_04	400001	510000
AE005174_05	500001	610000
AE005174_06	600001	710000
AE005174_07	700001	810000
AE005174_08	800001	910000
AE005174_09	900001	1010000
AE005174_10	1000001	1110000
AE005174_11	1100001	1210000
AE005174_12	1200001	1310000
AE005174_13	1300001	1410000
AE005174_14	1400001	1510000
AE005174_15	1500001	1610000
AE005174_16	1600001	1710000

```

AE005174_17 1700001 1810000
AE005174_18 1800001 1910000
AE005174_19 1900001 2010000
AE005174_20 2000001 2110000
AE005174_21 2100001 2210000
AE005174_22 2200001 2310000
AE005174_23 2300001 2410000
AE005174_24 2400001 2510000
AE005174_25 2500001 2610000
AE005174_26 2600001 2710000
AE005174_27 2700001 2810000
AE005174_28 2800001 2910000
AE005174_29 2900001 3010000
AE005174_30 3000001 3110000
AE005174_31 3100001 3210000
AE005174_32 3200001 3310000
AE005174_33 3300001 3410000
AE005174_34 3400001 3510000
AE005174_35 3500001 3610000
AE005174_36 3600001 3710000
AE005174_37 3700001 3810000
AE005174_38 3800001 3910000
AE005174_39 3900001 4010000
AE005174_40 4000001 4110000
AE005174_41 4100001 4210000
AE005174_42 4200001 4310000
AE005174_43 4300001 4410000
AE005174_44 4400001 4510000
AE005174_45 4500001 4610000
AE005174_46 4600001 4710000
AE005174_47 4700001 4810000
AE005174_48 4800001 4910000
AE005174_49 4900001 5010000
AE005174_50 5000001 5110000
AE005174_51 5100001 5210000
AE005174_52 5200001 5310000
AE005174_53 5300001 5410000
AE005174_54 5400001 5510000
AE005174_55 5500001 5528445
Continuation (40 of 56) of AE005174 from base 3900001 (AE005174 Escherichia coli O157:H7)

```

```

Query Match 96.8%; Score 1387.6; DB 1; Length 110000;
Best Local Similarity 98.0%; Pred. No. 0;
Matches 1405; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

Qy      1 ATGAAAGTAAAGCTGCGCAAGTTGAACGTGACGAGATGATGTGTGATGATG 60
Db      103596 ATGAAAGTAAAGCTGCGCAAGTTGAACGTGACGAGATGATGTGTGATGATG 103537

Qy      61 CTGGATCTGTAACGTGTAAGCGGCGCCACAGTCATCTCCCGGAAGCGCGGCGCTG 120
Db      103536 CTGGATCTGTAACGTGTAAGCGGCGCCACAGTCATCTCCCGGAAGCGCGGCGCTG 103477

Qy      121 GTTAAAGTAAATACATCGAAGACGTCCGCGCGCGGCTTAACGTGCGATGAATATC 180
Db      103476 GTTAAAGTAAATACATCGAAGACGTCCGCGCGCGGCTTAACGTGCGATGAATATC 103417

Qy      181 GCTTCTCTGCTATATGACGCTGTGTGCGGTTGACGGGCAATTGACGATGACGCGC 240
Db      103416 GCTTCTCTGCTATATGACGCTGTGTGCGGTTGACGGGCAATTGACGATGACGCGC 103357

Qy      241 GCGCTAGTAAATCTCTGCGCGGACGTCAAGTGAATGCACTTCGTTCTGTAACGAGC 300
Db      103356 GCGCTAGTAAATCTCTGCGCGGACGTCAAGTGAATGCACTTCGTTCTGTAACGAGC 103297

Qy      301 CATCCGACATTAACCAATTAACGGGTACTTCCCGCAACCAAGCTGATCCGCTGAT 360
Db      103296 CATCCGACATTAACCAATTAACGGGTACTTCCCGCAACCAAGCTGATCCGCTGAT 103237

Qy      361 TTGGAAGAGGTTTGAAGGTGATTCGCGACGCGCTGACGAGCGGATTAATCAAGCG 420
Db      103236 TTGGAAGAGGTTTGAAGGTGATTCGCGACGCGCTGACGAGCGGATTAATCAAGCG 103177

```

```

QY 421 CTGAGTTGATTCGCGCTGCTGCTTCTGACTACGCCAAGGTGCGTGGCAAGCTA 480
Db 103176 CTGAGTTGATTCGCGCTGCTGCTTCTGACTACGCCAAGGTGCGTGGCAAGCTA 103117
QY 481 CAGCAGATGATTCGCGCTGCTGCTTCTGACTACGCCAAGGTGCGTGGCAAGCTA 540
Db 103116 CAGCAGATGATTCGCGCTGCTGCTTCTGACTACGCCAAGGTGCGTGGCAAGCTA 103057
QY 541 ACCGATTTGAGCGCTACCGCGGCGCTACGCTGTTAAAGCGGATCTCTCGAATTTGAA 600
Db 103056 ACCGATTTGAGCGCTACCGCGGCGCTACGCTGTTAAAGCGGATCTCTCGAATTTGAA 102997
QY 601 GCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660
Db 102996 GCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 102937
QY 661 GCCGATTCGAACTCTGCGGCTCTGTTAGTACCCGTTCCGAAACGGGTATTCGCTGCTG 720
Db 102936 GCCGATTCGAACTCTGCGGCTCTGTTAGTACCCGTTCCGAAACGGGTATTCGCTGCTG 102877
QY 721 CAACCGGATTAAGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
Db 102876 CAACCGGATTAAGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 102817
QY 781 GGTGCGGCGCAACGCTGATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
Db 102816 GGTGCGGCGCAACGCTGATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 102757
QY 841 GAAGAAGCTGCTCTTCTTCTGCAATGCGCGGCTGCGTGTGTGTGTGTGTGTGTGTGT 900
Db 102756 GAAGAAGCTGCTCTTCTTCTGCAATGCGCGGCTGCGTGTGTGTGTGTGTGTGTGT 102697
QY 901 TCACGCTTTGCGCGATCGAGCTGAAATCTGTAGCTGAGCTGAGATACAGGCTTT 960
Db 102696 TCACGCTTTGCGCGATCGAGCTGAAATCTGTAGCTGAGCTGAGATACAGGCTTT 102637
QY 961 GCGCTGATGACCGAAGAGAACTGAAAGCTGCGCTGAGCGGCGCTGAACTGTGTGAA 1020
Db 102636 GCGCTGATGACCGAAGAGAACTGAAAGCTGCGCTGAGCGGCGCTGAACTGTGTGAA 102577
QY 1021 AAAGTGTGATGACCAACGCTGCTTCTGCAATCTGCAACGCGCGGCAAGCTCTTATCTG 1080
Db 102576 AAAGTGTGATGACCAACGCTGCTTCTGCAATCTGCAACGCGCGGCAAGCTCTTATCTG 102517
QY 1081 GCAATGCGCGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
Db 102516 GCAATGCGCGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 102457
QY 1141 AAACGCTGAAAGGGGATTCGCGCGCTGAAACCACTCGAAGCGCTATGATTCGCTG 1200
Db 102456 AAACGCTGAAAGGGGATTCGCGCGCTGAAACCACTCGAAGCGCTATGATTCGCTG 102397
QY 1201 GGGCGACTGGAAGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
Db 102396 GGGCGACTGGAAGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 102337
QY 1261 ATCCCGCGGATCTTTCGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
Db 102336 ATCCCGCGGATCTTTCGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 102277
QY 1321 GCCCGAGATTAAGAAAGTCTGCGGCAACGCTGCGCAAGTGTGTGTGTGTGTGTGTGTG 1380
Db 102276 GCCCGAGATTAAGAAAGTCTGCGGCAACGCTGCGCAAGTGTGTGTGTGTGTGTGTG 102217
QY 1381 GGTGCTGCAAGCAACATCATCAAGAGATTCGCAAGGATTAAGGCTAA 1434
Db 102216 GGTGCTGCAAGCAACATCATCAAGAGATTCGCAAGGATTAAGGCTAA 102163

```

RESULT 9  
 AB005174.40/c  
 WPCOMMENT  
 Sequence split into 56 fragments LOCUS AB005174 Accession AB005174

```

Fragment Name      Begin      End
AB005174.00        1          110000
AB005174.01        100001     210000
AB005174.02        200001     310000
AB005174.03        300001     410000
AB005174.04        400001     510000
AB005174.05        500001     610000
AB005174.06        600001     710000
AB005174.07        700001     810000
AB005174.08        800001     910000
AB005174.09        900001    1010000
AB005174.10       1000001    1110000
AB005174.11       1100001    1210000
AB005174.12       1200001    1310000
AB005174.13       1300001    1410000
AB005174.14       1400001    1510000
AB005174.15       1500001    1610000
AB005174.16       1600001    1710000
AB005174.17       1700001    1810000
AB005174.18       1800001    1910000
AB005174.19       1900001    2010000
AB005174.20       2000001    2110000
AB005174.21       2100001    2210000
AB005174.22       2200001    2310000
AB005174.23       2300001    2410000
AB005174.24       2400001    2510000
AB005174.25       2500001    2610000
AB005174.26       2600001    2710000
AB005174.27       2700001    2810000
AB005174.28       2800001    2910000
AB005174.29       2900001    3010000
AB005174.30       3000001    3110000
AB005174.31       3100001    3210000
AB005174.32       3200001    3310000
AB005174.33       3300001    3410000
AB005174.34       3400001    3510000
AB005174.35       3500001    3610000
AB005174.36       3600001    3710000
AB005174.37       3700001    3810000
AB005174.38       3800001    3910000
AB005174.39       3900001    4010000
AB005174.40       4000001    4110000
AB005174.41       4100001    4210000
AB005174.42       4200001    4310000
AB005174.43       4300001    4410000
AB005174.44       4400001    4510000
AB005174.45       4500001    4610000
AB005174.46       4600001    4710000
AB005174.47       4700001    4810000
AB005174.48       4800001    4910000
AB005174.49       4900001    5010000
AB005174.50       5000001    5110000
AB005174.51       5100001    5210000
AB005174.52       5200001    5310000
AB005174.53       5300001    5410000
AB005174.54       5400001    5510000
AB005174.55       5500001    5528445
Continuation (41 of 56) of AB005174 from base 4000001 (AB005174 Escherichia coli O157:H7)

Query Match      96.8%; Score 1387.6; DB 1; Length 110000;
Best Local Similarity 98.0%; Pred. No. 0;
Matches 1405; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

```

QY 1 ATGAAAGTAAAGCTGCGCAAGTTTGAAGTGCAGAGTGTGCTGCTGATGCTGATG 60  
 Db 3596 ATGAAAGTAAAGCTGCGCAAGTTTGAAGTGCAGAGTGTGCTGCTGATGCTGATG 3537  
 QY 61 CTGGAATGTTACTGTGTAAGCGCCACCAAGTCTATCTCGCCGGAAGCGCGTGCCTG 120  
 Db 3536 CTGGAATGTTACTGTGTAAGCGCCCTCAAGTCTATCTCGCCGGAAGCGCGTGCCTG 3477  
 QY 121 GTTAAAGTAAATCATCGAAGAAAGTCCGGGCGCGGCTTAAAGTGGGATGATATC 180

Db 3476 GTTAAAGTATACATCGAAGAAAGTCCGGGCGGCGGCTAAAGTGGCATGATATC 3417  
 Qy 181 GCTTCTCTGGTGTCTAATGACGCGCTGTGGGTGACGGGCATTTGACATGACGCGGC 240  
 Db 3416 GCTTCTCTGGTGTCTAATGACGCGCTGTGGGTGACGGGCATTTGACATGACGCGGC 3357  
 Qy 241 GCGGTGAGTAAATCTCTGGGCGGAGTCAAGTCAATGACGCTTGGTTTCTGTACGAGC 300  
 Db 3356 GCGGTGAGTAAATCTCTGGGCGGAGTCAAGTCAATGACGCTTGGTTTCTGTACGAGC 3297  
 Qy 301 CATCGACCATTTACCAATTACGGGTACTTTCGGCAACCAACAGTATCCGTCTGAT 360  
 Db 3296 CATCGACCATTTACCAATTACGGGTACTTTCGGCAACCAACAGTATCCGTCTGAT 3227  
 Qy 361 TTTGAAAGAGTTTCTGAAAGGTGTGATCCGACGCGCTGACGAGCGGATTAATCAGCG 420  
 Db 3236 TTTGAAAGAGTTTCTGAAAGGTGTGATCCGACGCGCTGACGAGCGGATTAATCAGCG 3177  
 Qy 421 CTGAGTTGATTTGGCGGCTGGTGTCTTCTGACTACGCAAGGATGGGCTGGCAAGGCTA 480  
 Db 3176 CTGAGTTGATTTGGCGGCTGGTGTCTTCTGACTACGCAAGGATGGGCTGGCAAGGCTA 3117  
 Qy 481 CAGCAGATGATCAATCTGCGCGCTGAAAGCGGCTTCCGCTGTGATTTGATCAAAAGT 540  
 Db 3116 CAGCAGATGATCAATCTGCGCGCTGAAAGCGGCTTCCGCTGTGATTTGATCAAAAGT 3057  
 Qy 541 ACCGATTTTGAAGCGCTACCGCGGCTACGCTGTTAACGCCGATCTCTCGAAATTTGAA 600  
 Db 3056 ACCGATTTTGAAGCGCTACCGCGGCTACGCTGTTAACGCCGATCTCTCGAAATTTGAA 2997  
 Qy 601 GCTGTGTGCGTAAATGATGAAACGGAAGAGATTTGTGAGCGCGGATGAACTGAT 660  
 Db 2996 GCTGTGTGCGTAAATGATGAAACGGAAGAGATTTGTGAGCGCGGATGAACTGAT 2937  
 Qy 661 GCCGATTACGAATCTCGGCTCTGTGATGACCGCTTCCGAACAGGATGTCGCTG 720  
 Db 2936 GCCGATTACGAATCTCGGCTCTGTGATGACCGCTTCCGAACAGGATGTCGCTG 2877  
 Qy 721 CAACCGGCTAAAGCGCGCTGATATGCAACCAAGCGCAGAAAGTATGACGTTACC 780  
 Db 2876 CAACCGGCTAAAGCGCGCTGATATGCAACCAAGCGCAGAAAGTATGACGTTACC 2817  
 Qy 781 GGTGCGGCGGACGACGGTATTTGGGCTCTGGCGGCAACGCTGGAGGGGTAATTCGCTG 840  
 Db 2816 GGTGCGGCGGACGACGGTATTTGGGCTCTGGCGGCAACGCTGGAGGGGTAATTCGCTG 2757  
 Qy 841 GAAGAAGCTGCTCTTTGCAATGCGGCGCTGCTGTGTGTCTGCGCAACCTGGAAAC 900  
 Db 2756 GAAGAAGCTGCTCTTTGCAATGCGGCGCTGCTGTGTGTGTCTGCGCAACCTGGAAAC 2697  
 Qy 901 TCCACGGTTTTCGCGATGAGCTGAAATGCTGTGATGCTGACATGACAGGCTTT 960  
 Db 2696 TCCACGGTTTTCGCGATGAGCTGAAATGCTGTGATGCTGACATGACAGGCTTT 2637  
 Qy 961 GGGGTGATGACGAAGGAAGTGAAGTGGCGGTAGCGGAGGGCGGTAACGTTGAA 1020  
 Db 2636 GGGGTGATGACGAAGGAAGTGAAGTGGCGGTAGCGGAGGGCGGTAACGTTGAA 2577  
 Qy 1021 AAAGTGTGATGACCAACGCTGTCTTGAATCTGTGACGCGCGGACGCTCTTATCTG 1080  
 Db 2576 AAAGTGTGATGACCAACGCTGTCTTGAATCTGTGACGCGCGGACGCTCTTATCTG 2517  
 Qy 1081 GCAAAATGCGCGAAGCTGGGTGACCGCTTGAATTTGTCGCTGCAACAGGATGCTTCAAC 1140  
 Db 2516 GCAAAATGCGCGAAGCTGGGTGACCGCTTGAATTTGTCGCTGCAACAGGATGCTTCAAC 2457  
 Qy 1141 AAAGGCTGAAAGGGGATTTCCGCGCGGTAACCACTCGAACAGCGTATGTTGTG 1200  
 Db 2456 AAAGGCTGAAAGGGGATTTCCGCGCGGTAACCACTCGAACAGCGTATGTTGTG 2397  
 Qy 1201 GGGCAGCTGAAAGCGCTGACTGGGTAGTGTGTTTGAAGAGACACGCGCAGCGCTTG 1260  
 Db 2396 GGGCAGCTGAAAGCGCTGACTGGGTAGTGTGTTTGAAGAGACACGCGCAGCGCTTG 2337

Qy 1261 ATGCGCGGATCTTTCGACATCTGCTGTGAAGGCGCGGACATTAACAGAGAT 1320  
 Db 2336 ATGCGCGGATCTTTCGACATCTGCTGTGAAGGCGCGGACATTAACAGAGAT 2277  
 Qy 1321 GCGGAGTAAAGTCTGGGCAACGCTGGGAGTGTGGTCTCACTTTGAAGAC 1380  
 Db 2276 GCGGAGTAAAGTCTGGGCAACGCTGGGAGTGTGGTCTCACTTTGAAGAC 2217  
 Qy 1381 GGTGCTGACGACCAATCATGAGAGATCAACAGATTAAGGCTTA 1434  
 Db 2216 GGTGCTGACGACCAATCATGAGAGATCAACAGATTAAGGCTTA 2163

RESULT 10  
 BA000007\_39/c  
 WPCOMMENT  
 Sequence split into 55 fragments Locus BA000007 Accession BA000007

Fragment Name	Begin	End
BA000007_00	1	11000
BA000007_01	100001	21000
BA000007_02	200001	31000
BA000007_03	300001	41000
BA000007_04	400001	51000
BA000007_05	500001	61000
BA000007_06	600001	71000
BA000007_07	700001	81000
BA000007_08	800001	91000
BA000007_09	900001	101000
BA000007_10	1000001	111000
BA000007_11	1100001	121000
BA000007_12	1200001	131000
BA000007_13	1300001	141000
BA000007_14	1400001	151000
BA000007_15	1500001	161000
BA000007_16	1600001	171000
BA000007_17	1700001	181000
BA000007_18	1800001	191000
BA000007_19	1900001	201000
BA000007_20	2000001	211000
BA000007_21	2100001	221000
BA000007_22	2200001	231000
BA000007_23	2300001	241000
BA000007_24	2400001	251000
BA000007_25	2500001	261000
BA000007_26	2600001	271000
BA000007_27	2700001	281000
BA000007_28	2800001	291000
BA000007_29	2900001	301000
BA000007_30	3000001	311000
BA000007_31	3100001	321000
BA000007_32	3200001	331000
BA000007_33	3300001	341000
BA000007_34	3400001	351000
BA000007_35	3500001	361000
BA000007_36	3600001	371000
BA000007_37	3700001	381000
BA000007_38	3800001	391000
BA000007_39	3900001	401000
BA000007_40	4000001	411000
BA000007_41	4100001	421000
BA000007_42	4200001	431000
BA000007_43	4300001	441000
BA000007_44	4400001	451000
BA000007_45	4500001	461000
BA000007_46	4600001	471000
BA000007_47	4700001	481000
BA000007_48	4800001	491000
BA000007_49	4900001	501000
BA000007_50	5000001	511000
BA000007_51	5100001	521000
BA000007_52	5200001	531000
BA000007_53	5300001	541000

BA000007.54 5400001 5498450  
 Continuation (40 of 55) of BA000007 from base 3900001 (BA000007 Escherichia coli O157:H7)

Query Match 96.8%; Score 1387.6; DB 1; Length 110000;  
 Best Local Similarity 98.0%; Pred. No. 0;  
 Matches 1405; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

1 ATGAAAGTACGCTGCGCAGATTTGAACGTGACAGAGTATGTGTGTGTGTATG 60  
 Db ATGAAAGTACGCTGCGCAGATTTGAACGTGACAGAGTATGTGTGTGTGTATG 36222  
 Qy CTGATCTGTTACGTGATCGGCGCCCAACGATCTCTGCGGAAAGCGCCGCTG 120  
 Db CTGATCTGTTACGTGATCGGCGCCCAACGATCTCTGCGGAAAGCGCCGCTG 36222  
 Qy 121 GTTAAAGTATACCATGGAAGAAAGTCCGGCGCGCGCTTACGTGGCGATATATC 180  
 Db GTTAAAGTATACCATGGAAGAAAGTCCGGCGCGCGCTTACGTGGCGATATATC 36102  
 Qy 181 GCTTCTCTGCGTGAATGACAGCGCTGGTGGGTTGACGGGCAATTGACGATGACGCGC 240  
 Db GCTTCTCTGCGTGAATGACAGCGCTGGTGGGTTGACGGGCAATTGACGATGACGCGC 36042  
 Qy 241 GCGGTGAATTAATCTCTGCGCGACGTCACGTCATTAATGCACTTCTGTTCTGTA 300  
 Db GCGGTGAATTAATCTCTGCGCGACGTCACGTCATTAATGCACTTCTGTTCTGTA 35982  
 Qy 301 CATCCGACATTACCAATTAACGGGTACTTCCCGCAACCAACGATGATCCGCTGAT 360  
 Db CATCCGACATTACCAATTAACGGGTACTTCCCGCAACCAACGATGATCCGCTGAT 35922  
 Qy 35981 CATCCGACATTACCAATTAACGGGTACTTCCCGCAACCAACGATGATCCGCTGAT 35922  
 Db TTTGAAGAGGTTTGAAGAGTGTGATCCGACGCGCTGACGAGCGATTAATCAGCG 420  
 Qy 35221 TTTGAAGAGGTTTGAAGAGTGTGATCCGACGCGCTGACGAGCGATTAATCAGCG 35862  
 Db CTGAGTTGATGCGCGCTGCTGCTTTCTGACTACGCAAAAGTGGCGTGAAGCGTA 480  
 Qy 35861 CTGAGTTGATGCGCGCTGCTGCTTTCTGACTACGCAAAAGTGGCGTGAAGCGTA 35802  
 Db CAGCAGATGATCCCACTGCGCGCTGTAAGCGGGTGTTCGGGTGATGATTAACCAAAAGT 540  
 Qy 481 CAGCAGATGATCCCACTGCGCGCTGTAAGCGGGTGTTCGGGTGATGATTAACCAAAAGT 540  
 Db CAGCAGATGATCCCACTGCGCGCTGTAAGCGGGTGTTCGGGTGATGATTAACCAAAAGT 35742  
 Qy 541 ACCGATTTGAGCGCTACCGCGCGCTACGCTGTTTAAACCGCAATCTCTCGAATTTGAA 600  
 Db ACCGATTTGAGCGCTACCGCGCGCTACGCTGTTTAAACCGCAATCTCTCGAATTTGAA 35741  
 Qy 601 GCTGTTGCGGTAAATGTAAGACCGAAGAGAGATTTGAGCGCGGATGAACTGAT 660  
 Db GCTGTTGCGGTAAATGTAAGACCGAAGAGAGATTTGAGCGCGGATGAACTGAT 35681  
 Qy 661 GCCGATTTACGAATCTGCGGCTCTGTTAGTGAACCGGTCGCAAGGGATTAATCCCTGCTG 720  
 Db GCCGATTTACGAATCTGCGGCTCTGTTAGTGAACCGGTCGCAAGGGATTAATCCCTGCTG 35621  
 Qy 721 CAACCGGGTAAAGCGCGCTCATATGCAACCAACGAGCGAGAGTATGAGTTAAC 780  
 Db CAACCGGGTAAAGCGCGCTCATATGCAACCAACGAGCGAGAGTATGAGTTAAC 35561  
 Qy 781 GGTGCGGGCGACACGCGTATTTGCGTCTGCGCGCAACGCTGGCAGCGGGTAAATTCGCTG 840  
 Db GGTGCGGGCGACACGCGTATTTGCGTCTGCGCGCAACGCTGGCAGCGGGTAAATTCGCTG 35501  
 Qy 841 GAAAGAGCTGCTCTTTTCCCAATGCGCGCGCTGCGGTGCTGCGCAATCTGGAAC 900  
 Db GAAAGAGCTGCTCTTTTCCCAATGCGCGCGCTGCGGTGCTGCGCAATCTGGAAC 35441  
 Qy 901 TCCACGCTTTGCGCGATCGAGCTGGAATAATCTGTACGTGACGTGACGATACAGCTTT 960  
 Db TCCACGCTTTGCGCGATCGAGCTGGAATAATCTGTACGTGACGTGACGATACAGCTTT 35381  
 Qy 961 GCGGTGATGACCGAAGAGAACTGAAGCTGCGGTAGCGGAGCGCGTAAACGTGTGAA 1020  
 Db GCGGTGATGACCGAAGAGAACTGAAGCTGCGGTAGCGGAGCGCGTAAACGTGTGAA 35322

Db 35321 GCGGTGATGACCGAAGAGAACTGAAGCTGCTGTATGAGCGAGCGGTAAACGTGTGAA 35262  
 Qy 1021 AAGGTGATGATGACCAACGGTGTCTTTGACATCTGTCACGCGGGAGACGTCTTATCTG 1080  
 Db AAGGTGATGATGACCAACGGTGTCTTTGACATCTGTCACGCGGGAGACGTCTTATCTG 35261  
 Qy 1081 GCAAATGCCCGACAGCTGCGTGAACCGCTTATGTTGTGCGGTCAACAGCATGCTCCACC 1140  
 Db GCAAATGCCCGACAGCTGCGTGAACCGCTTATGTTGTGCGGTCAACAGCATGCTCCACC 35201  
 Qy 1141 AAACGCTGAAGAGGATTTCCCGCGCTGAACCCACTGCAACAGCGTATGTTGCTG 1200  
 Db AAACGCTGAAGAGGATTTCCCGCGCTGAACCCACTGCAACAGCGTATGTTGCTG 35141  
 Qy 1201 GCGGCACTGGAACGCGTGAACGCTGAGTGTGTTGAAGAGGACAGCGCGAGCGCTG 1260  
 Db GCGGCACTGGAACGCGTGAACGCTGAGTGTGTTGAAGAGGACAGCGCGAGCGCTG 35081  
 Qy 1261 ATGCGCGGATCTTTCGACATCTGCTGTGAAGGCGCGCATTAACCAAGAGATT 1320  
 Db ATGCGCGGATCTTTCGACATCTGCTGTGAAGGCGCGCATTAACCAAGAGATT 35021  
 Qy 1321 GCCGGAGTAAAGAGTCTGGCGCAACGCGTGGCAAGTGTGTGCTCACTTTGAAGAC 1380  
 Db GCCGGAGTAAAGAGTCTGGCGCAACGCGTGGCAAGTGTGTGCTCACTTTGAAGAC 34961  
 Qy 1381 GGTGTGCGACGACCAACATCATCAAGAGATCCACAGGATTAATAAGGCTAA 1434  
 Db GGTGTGCGACGACCAACATCATCAAGAGATCCACAGGATTAATAAGGCTAA 34901

RESULT 11  
 AE016767/c 300359 bp DNA linear BCT 24-DEC-2002  
 LOCUS  
 DEFINITION Escherichia coli CFT073 section 13 of 18 of the complete genome.  
 ACCESSION AE016767 AE014075  
 VERSION AE016767.1 GI:26110054  
 KEYWORDS  
 SOURCE  
 ORGANISM Escherichia coli CFT073  
 Escherichia coli CFT073  
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 Enterobacteriaceae; Escherichia.

REFERENCE  
 AUTHORS  
 1 (bases 1 to 300359)  
 Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roessch,P.,  
 Resko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,D.,  
 Stroud,P., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,  
 Perna,N.T., Mobley,H.L.T., Domeneberg,M.S., and Blattner,F.R.  
 Extensive Mosiac Structure Revealed by the Complete Genome Sequence  
 of Uropathogenic Escherichia coli  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 17020-17024 (2002)  
 12471157

TITLE  
 JOURNAL  
 PUBMED  
 AUTHORS  
 TITLE  
 JOURNAL  
 SUBMITTED (20-JUN-2002) Genetics Laboratory, University of  
 Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA  
 Location/Qualifiers  
 1. 300359

FEATURES  
 source  
 /organism="Escherichia coli CFT073"  
 /mol\_type="genomic DNA"  
 /strain="CFT073"  
 /db\_xref="taxon:199310"  
 188..547  
 /gene="yqjC"  
 /locus\_tag="C3796"  
 188..547  
 /gene="yqjC"  
 /locus\_tag="C3796"  
 /note="Escherichia coli K-12 ortholog: b3042; Escherichia  
 coli O157:H7 ortholog: z4400"

CDS  
 188..547  
 /locus\_tag="C3796"  
 /note="Escherichia coli K-12 ortholog: b3042; Escherichia  
 coli O157:H7 ortholog: z4400"

gene  
/codon\_start=1  
/transl\_table=11  
/product="Hypothetical protein yqkC"  
/protein\_id="AN82241.1"  
/db\_xref="GI:26110055"  
/translation="MTYPRRTMALNTYYQRETTMDPKKIBQIARQVHSMKPIRE  
FGDVEKKIRQTLQAOLTRDLVSRBSFDVQTVLALTRKMLLEGRISLELRVNP  
VALQSPRAPPPDKAE"  
complement(582..788)  
/gene="g1gS"  
/locus\_tag="c3797"  
complement(582..788)  
/gene="g1gS"  
/locus\_tag="c3797"  
/functions="putative enzyme; Macromolecule synthesis,  
modification; Polysaccharides - (cytoplasmic)"  
/note="Escherichia coli K-12 ortholog: b3049; Escherichia  
coli O157:H7 ortholog: z4401"  
/codon\_start=1  
/transl\_table=11  
/product="Glycogen synthesis protein g1gS"  
/protein\_id="AN82242.1"  
/db\_xref="GI:26110056"  
/translation="MLMDHSLNSLNFDFLARSFARMHAEGRPVDLIAYTGNDDEHR  
TWCFARYAWYCCQWMTRELELEH"  
1052..1681  
/gene="yqkU"  
/locus\_tag="c3798"  
1052..1681  
/gene="yqkU"  
/locus\_tag="c3798"  
/functions="putative"  
/note="Escherichia coli K-12 ortholog: b3050; Escherichia  
coli O157:H7 ortholog: z4402"  
/codon\_start=1  
/transl\_table=11  
/product="Hypothetical protein yqkU"  
/protein\_id="AN82243.1"  
/db\_xref="GI:26110057"  
/translation="MILFADYNTPEYLFALISFVLLIGLLEIFALICGMLSGALDAHID  
HYDSITTHGSIQALMYLNIQRLPALVVLCSLAFPGILQIQAQVTLWQSPPLNF  
VPSLPLFTIIVAYTGKVVAPMIPRSHSAITEEBYSGMALITGHQATSGNPECG  
LTQPFQIHYLLLEPEBEGKTFTKDKVLIICRLSATRYLLENNPWSQIL"  
1708..3369  
/gene="yqkK"  
/locus\_tag="c3799"  
1708..3369  
/gene="yqkK"  
/locus\_tag="c3799"  
/functions="putative membrane"  
/note="Escherichia coli K-12 ortholog: b3051; Escherichia  
coli O157:H7 ortholog: z4403"  
/codon\_start=1  
/transl\_table=11  
/product="Hypothetical protein yqkK"  
/protein\_id="AN82244.1"  
/db\_xref="GI:26110058"  
/translation="WDDIVNSVPSWMTAIIAVCIIFIGIIPARLYRRASQAQVFR  
TGLSGQVWVSGAIVMPIFHBIIPIMNTLKLVSRTSDISLITKDRMVDVVAF  
VRVPSVEGIATAQTLGORTLSPEIDRLMVEDFVADALRAQMTMHLOTRERF  
VOGVONTVAEDLSKNGLESVSLTNFNQSKHFNPNNAFDABGLTKLQOERRRR  
ERRNEVDVAVAREKRDALSRLETLEOBAFPTLEBOOVTRTAEONAKIAPPA  
ERRRRAQTRILARQIQETETIDREQAVRSKVEAEKRVIKIIEQOQVETIANQTS  
ITALNAEQSQSQAERANLALAEVSAQNVETTRQAEADRAQVALIAADDAETK  
FALITVALAKAEKAEWQAAIYELAEATRKGLAEARALANLAINVLSIDQTSIK  
FKALITQALPAVIEKSVPMKAIIDGIIKIIVDGLNRGATGADANTGVGGANLAEQL  
SALSVRTQAPLIDSLNBIIGVSGGSLAAL7TSLSSTTPVEKAE"  
complement(3788..5221)  
/locus\_tag="c3800"  
complement(3788..5221)  
/locus\_tag="c3800"  
/BC\_number="2.7.-.-"  
/function="putative"

gene  
/note="Escherichia coli K-12 ortholog: b3052; Escherichia  
coli O157:H7 ortholog: z4405"  
/codon\_start=1  
/transl\_table=11  
/product="ADP-heptose synthase"  
/protein\_id="AN82245.1"  
/db\_xref="GI:26110059"  
/translation="MKVTLPEPERAGVWVGVDMLDRLVWYGPTRISPEPVPVKN  
TIERPGAAVNAVNIASIGANRLVGLTGIDDAARALSKSLADVNVKDFVSPVPH  
TITKLVLNRNQLRLDPFEGFEGVDPPQLERINQALSSIGALITSDYAKALASV  
QQM1QLARAGAPVLLIDPGTDERYRAGTLLTPNLSFEPAVVGKCTEBEIVERMK  
LIADYELSLVTRSEQMSLLQPGKAPLMPQOEVVDVYAGDVTGVGLAATLAA  
GNSLEBAECPANAAAGVWVKLGSTVSPILENARVGRADTFGVTSEELATAVA  
ARRGRKVMVNTGVFDIILHAGHYLANARKGDLRLIVVNSPASTKRLKGDSPVNP  
LEQPMVLCALREAVPVVSPEBTPPRLAGILPDLVYNGDVKEBISKEMVANG  
GEVLVNFEGCGSTNIIKIQDKRG"  
complement(5269..8220)  
/gene="g1nB"  
/locus\_tag="c3801"  
complement(5269..8220)  
/gene="g1nB"  
/locus\_tag="c3801"  
/BC\_number="2.7.7.42"  
/function="enzyme; Proteins - translation and  
modification"  
/note="Escherichia coli K-12 ortholog: b3053; Escherichia  
coli O157:H7 ortholog: z4406"  
/codon\_start=1  
/transl\_table=11  
/product="Glutamate-ammonia-lyase adenylyltransferase"  
/protein\_id="AN82246.1"  
/db\_xref="GI:26110060"  
/translation="MRGNATKINISVMOGTIRNSGCTAENVNOARIKSIMKLSPP  
LQOYVTVVARELPLEBSLSAQASVTLFSDFPVDSVTAPHEMTLEBSQPDQE  
MOWYASWMLQELANSVSDSEAGLMEKRLFRRLINVRILAMQTLALVTBESLLOQLSHA  
ETLIVARMDLYDACCEWGTPCNAQGEAPVLLIDMGKLGGLNLSFDILIFAMP  
BHCCTCGGRRELDNAPFTFMQGRLLIKVDPQDPDQVVRVMDRLRPFSGSPLVFSF  
ALLEDYVOGGRMERVYAWYKARIMGDSAGTAVNELRAMLRPVFRFYIDFVYQSLR  
NMKGMIAREYRRRGITDNIKLRAGGIRREIYVQVQOLIRGSEBPSLSALLPTLSA  
IAPLHLSENDABQLAVATILFRRLNLSQINDBOTQTLPRBELARALAMMDPAD  
MPODTQVLAHMANVRVNEILIGDESSTOBSLSQREMLQDQEDVTPVLAH  
LESDRTOVLAHMANVRVNEILIGDESSTOBSLSQREMLQDQEDVTPVLAH  
ALVGVITRTYVLELSRFPALKHLISLCAASPMASOLARLPLDLDELDPNTLYQ  
PATDVAVRBELRGVILRVPEDEDEQOLISARQKQOLIRIAADTAGTLVWVKVSDH  
LTYLARAMDVAYQAAVQWVYARXGPNHLEBEGRGAVNGYKCGEELGVSSDLD  
LIFHDCEPMDATDGERBIDGROFYLRQRLIMHLSSTSSGILYEVDAKLAPSGAA  
GMLVTSAEAFADYQKNBAMTWERQALVRVAVGDPQLTAHPFAVAREMTLPRBGT  
LQTEVEREMERKMAHLGNKHRRFDIKADEGGITDIEFTQVYLVAREHKEPGLTWS  
DNVRLLELAONDIMEQESAMALTRAAYTTLRDELHMLALQELRGVSBDCFALREILV  
RASQWQWIVBE"  
complement(8132..9460)  
/gene="yqkF"  
/locus\_tag="c3802"  
complement(8132..9460)  
/gene="yqkF"  
/locus\_tag="c3802"  
/note="TORXZ; Escherichia coli K-12 ortholog: b3054;  
Escherichia coli O157:H7 ortholog: z4407"  
/codon\_start=1  
/transl\_table=11  
/product="Hypothetical protein yqkF"  
/protein\_id="AN82247.1"  
/db\_xref="GI:26110061"  
/translation="WGVPSGGMAQIEIKPTVNSHVAEALRDHLNTLOGEHHNDVQ  
LNIIVTEPDNMPNGLGDMGRLRGSEGRREMTKVAQVGTGLHQEPENVALSEPTL  
DLAQLPTEWPNELPADLASRVQPLFSDPFYBEKMLVEVDGSRIRIADPOGVKGE  
FAPPICELEIELSGMRAVILKLANQLVSGTGRQSLSVDAAGVLAAGNPAREIKP  
TTIIVAAKADVEQGLBALLELAQMOVHEBLMWGNDAAKQVAAAGLVHTIIML  
FGGIVPRKASTHLDLLTOCEATIASAVSAVAVVSTTRMAKLTALTEMLVSAKQBP  
LDKKAOSKISDSFTRPADTHLSRHAELSVFQPIGDRRDDLPALITRDISILLLA  
GVYDPAQDQWLENWQLSRHAIATGQRIEIBHRNANNOBPPWLHSGRK"  
complement(9503..9955)  
/locus\_tag="c3803"

CDS  
complement (9503.. .9955)  
/locus\_tag="C3803"  
/function="putative enzyme: Transposases"  
/note="Escherichia coli O157:H7 ortholog: 25815"  
/codon\_start=1  
/transl\_table=1  
/product="Putative Transposase"  
/protein\_id="AA082248.1"  
/db\_xref="GI:26110062"  
/translation="MWLFPMNTSTFFMEQGYCEINRSYSAVLHVVHVPKTKR  
PVEGLHANNRQYVAECADPGAELECEGSEADHVMLEPPVLTPLVNSLKS  
TSRLRNDPFLDRAAYSKPVLMSRYPYIGSGGAPLEVVKKYIQNDRG"  
9968..11173  
/gene="ydcW"  
Query Match 96.3%; Score 1381.2; DB 1; Length 300359;  
Blast Local Similarity 97.7%; Pred. No. 2.2e-312;  
Matches 1401; Conservative 0; Mismatches 33; Indels 0; Gaps 0;  
QY 1 ATGAAAGTAAAGCTGCGCAGATTTGAACGTGCAAGAGTGTGTGTGATGTGATG 60  
DB 5221 ATGAAAGTAAAGCTGCGCAGATTTGAACGTGCAAGAGTGTGTGTGATGTGATG 5162  
QY 61 CTGGATCGTTACTGTGTAAGCGCCCAACCAATCTTCCGGAAGCGCGGTCCTG 120  
DB 5161 CTGGATCGTTACTGTGTAAGCGCCCAACCAATCTTCCGGAAGCGCGGTCCTG 5102  
QY 121 GTTAAAGTAAATACATCTGAAGAAAGTCCGGGCGCGGCTAATCGTGGCAATGAATTC 180  
DB 5101 GTTAAAGTAAATACATCTGAAGAAAGTCCGGGCGCGGCTAATCGTGGCAATGAATTC 5042  
QY 181 GCTTCTCTCGGTGCTAATGACGCTGGTGGGTTGAAGGAGTGAAGATGACAGCGCG 240  
DB 5041 GCTTCTCTCGGTGCTAATGACGCTGGTGGGTTGAAGGAGTGAAGATGACAGCGCG 4982  
QY 241 GCGGTGATTAATCTTGGCGGACGTCACGTCGAATGCGATTCCTTCTGTAACGACG 300  
DB 4981 GCGGTGATTAATCTTGGCGGACGTCACGTCGAATGCGATTCCTTCTGTAACGACG 4922  
QY 301 CATCGGACCATTAACCAATTAACGGGTACTTCCCGCAACCAAGCTGATCCCGTGGAT 360  
DB 4921 CATCGGACCATTAACCAATTAACGGGTACTTCCCGCAACCAAGCTGATCCCGTGGAT 4862  
QY 361 TTGGAAGAGGTTTGAAGGTGTGATCCGACGCGCTGACAGCGGATTAATCAGCGC 420  
DB 4861 TTGGAAGAGGTTTGAAGGTGTGATCCGACGCGCTGACAGCGGATTAATCAGCGC 4802  
QY 421 CTGAGTTGATTTGGCGGCTGTGCTTTCTGATACGCGCAAAAGTGGCGTGGCAAGCGTA 480  
DB 4801 CTGAGTTGATTTGGCGGCTGTGCTTTCTGATACGCGCAAAAGTGGCGTGGCAAGCGTA 4742  
QY 481 CAGAGATGATCCAACTGGCGCGTGAAGCGGGGTTCGCGGTGATGATTAACCAAAAGT 540  
DB 4741 CAGAGATGATCCAACTGGCGCGTGAAGCGGGGTTCGCGGTGATGATTAACCAAAAGT 4682  
QY 541 ACCGATTTTGAAGCGCTACCGCGCGCTGATCGCTTTAAACCGCAATCTCTCGAATTTGAA 600  
DB 4681 ACCGATTTTGAAGCGCTACCGCGCGCTGATCGCTTTAAACCGCAATCTCTCGAATTTGAA 4622  
QY 601 GCTGTGTGCTGTAATGTAAGACCGAAGAGATTTGTAAGCGCGCGCATGAATCTGAT 660  
DB 4621 GCTGTGTGCTGTAATGTAAGACCGAAGAGATTTGTAAGCGCGCGCATGAATCTGAT 4562  
QY 661 GCCGATTAACGAACCTCGGCTCTGTGTGTGACCGGTTCCGAACAGGGGATGTCCTGCTG 720  
DB 4561 GCCGATTAACGAACCTCGGCTCTGTGTGTGACCGGTTCCGAACAGGGGATGTCCTGCTG 4502  
QY 721 CAACCGGGTAAAGCGCGCTCATATGCAACCAAGCGCAGAAAGTATGATGATTAAC 780  
DB 4501 CAACCGGGTAAAGCGCGCTCATATGCAACCAAGCGCAGAAAGTATGATGATTAAC 4442  
QY 781 GGTGCGGGCGACACGGTATTTGGCGTCTGTGCGCGCAACGCTGGACGCGGATTAATTCGCTG 840

DB 4441 GGTGCGGGCGACACGGTATTTGGCGTCTGTGCGCGCAACCGTGGACGCGGATTAATTCGCTG 4382  
QY 841 GAGAAAGCGTCTCTTTGCAATGCGGGCGCTGCGTGGTGTGCGCAAACTGGGAAC 900  
DB 4381 GAGAAAGCGTCTCTTTGCAATGCGGGCGCTGCGTGGTGTGCGCAAACTGGGAAC 4322  
QY 901 TCCACGTTTCCGCGATCGAGCTGGAATAATGCTGTGACGTGACGTGCAATACAGGCTTT 960  
DB 4321 TCCACGTTTCCGCGATCGAGCTGGAATAATGCTGTGACGTGACGTGCAATACAGGCTTT 4262  
QY 961 GCGGTGATGACCGAAGAGAACTGAAGCTGCGCGTGGCGGCGGTAATGCTGGAA 1020  
DB 4261 GCGGTGATGACCGAAGAGAACTGAAGCTGCGCGTGGCGGCGGTAATGCTGGAA 4202  
QY 1021 AAGGTGATGACCAACCGTGTCTTGAATCTGACATCTGACACGCGGGACGCTCTTAATCTG 1080  
DB 4201 AAGGTGATGACCAACCGTGTCTTGAATCTGACATCTGACACGCGGGACGCTCTTAATCTG 4142  
QY 1081 GCMAATGCCCGCAAGCTGGTGAACGCTGATTTGTCGCTCAACAGCGATGCTCCACC 1140  
DB 4141 GCMAATGCCCGTAACTGGGTGAACGCTGATTTGTCGCTGTAACAGCGATGCTCCACC 4082  
QY 1141 AAGCGCTGAAGAGGGAATTCGCGCGCGGTAAACCACTGGAACAGGCTAATGATGCTG 1200  
DB 4081 AAGCGCTGAAGAGGGAATTCGCGCGCGGTAAACCACTGGAACAGGCTAATGATGCTG 4022  
QY 1201 GCGCACTGGAACCGGTGCACTGGGTAGTGTGCTTTGAAGAGGACACGCGGACGCTTG 1260  
DB 4021 GCGCACTGGAACCGGTGCACTGGGTAGTGTGCTTTGAAGAGGACACGCGGACGCTTG 3962  
QY 1261 ATGCGCGGATCTTGGCAGATCTGCTGTGAAAGCGCGGCACTAATAACCAAGAGATT 1320  
DB 3961 ATGCGCGGATCTTGGCAGATCTGCTGTGAAAGCGCGGCACTAATAACCAAGAGATT 3902  
QY 1321 GCGCGGATTAAGAAAGTCTGGGCAACGCTGGGGAAGTGTGCTGCTCACTTTGAAGAC 1380  
DB 3901 GCGCGGATTAAGAAAGTCTGGGCAACGCTGGGGAAGTGTGCTGCTCACTTTGAAGAC 3842  
QY 1381 GGTGCTGCGACGCAACATCATCAAGAGATCAACAGGATTAATAAGGCTAA 1434  
DB 3841 GGTGCTGCGACGCAACATCATCAAGAGATCAACAGGATTAATAAGGCTAA 3788  
  
RESULT 12  
AE016988/c  
LOCUS  
DEFINITION  
Shigella flexneri 2a str. 2457T section 11 of 16 of the complete  
genome.  
ACCESSION  
AE016988 AE014073  
VERSION  
AE016988.1 GI:30042420  
KEYWORDS  
SOURCE  
ORGANISM  
Shigella flexneri 2a str. 2457T  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Shigella.  
REFERENCE  
1 (bases 1 to 292906)  
Wei, J., Goldberg, M. B., Burland, V., Venkatesan, M. M., Deng, W.,  
Fournier, G., Mayhew, G. F., Plunkett, G. III, Rose, D. J., Darling, A.,  
Mau, B., Perna, N. T., Payne, S. M., Runyen-Janecky, L. J., Zhou, S.,  
Schwartz, D. C., and Blattner, F. R.  
Complete Genome Sequence and Comparative Genomics of Shigella  
flexneri Serotype 2a Strain 2457T  
Infect. Immun. 71 (5), 2775-2786 (2003)  
12704152  
JOURNAL  
PUBMED  
REFERENCE  
2 (bases 1 to 292906)  
Wei, J., Goldberg, M. B., Burland, V., Venkatesan, M. M., Deng, W.,  
Fournier, G., Mayhew, G. F., Plunkett, G. III, Rose, D. J., Darling, A.,  
Mau, B., Perna, N. T., Payne, S. M., Runyen-Janecky, L. J., Zhou, S.,  
Schwartz, D. C., and Blattner, F. R.  
Submitted (13-JUN-2002) Genetics Laboratory, University of  
Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA  
location/Qualifiers  
FEATURES





## CDS

/locus\_tag="B3037"  
 complement(7274..9520)  
 /gene="ptab"  
 /locus\_tag="B3037"  
 /function="enzyme; Transport of small molecules: Other"  
 /note="Residues 1 to 736 of 748 are 99.59 pct identical to  
 residues 1 to 736 of 748 from Escherichia coli K-12 ;  
 B2829"  
 /codon\_start=1  
 /transl\_table=1  
 /product="PTS system, enzyme I, transcriptional regulator  
 (with Npr and Npr proteins)"  
 /protein\_id="AAP1815.1"  
 /db\_xref="gi:30042427"  
 /translation="MLTRIREIVKVASPAPINBALNIVTDICLMDTEVCVYLLD  
 HDRCYILMTNRGAKKRRGRVTYLAPEBGIVGLYGRALPEINLADAKHSPKIPESV  
 KEERFRAVLGPVLIQRQLGVLVQDRELMQDSESESPVLTATMAILSOSLT  
 ALFGQYQYRIRALPAAPGVAIAGMODATLPLMEQYQASITLDPALERRLTGLER  
 AANFRYSKRFAPGAQKETAIFDLVSHLSIDRLRELAEVDKGSVMAVATVI  
 EKFAEPALSDNYLKERAGDRLALGOLRFLHLDANQNPAPERFLVADLSATT  
 LAELPDRIVGVVNRDGAANSHALIMRALGIPVMGADIDPSVLRHTLIDVYRGE  
 LIDPEPVTLQRYRLISERIEISRLIABDDNLEPAQKSGERITVMLNAGLSPEHEK  
 LGSCIDIGLRTIRIPMLQSPFSESVQAQVQMLQMFENDKPVTLITLDVAGKOL  
 PYMPISENPCLGWRGIRITLDQEIFLIQVRAMLRAANAATGNINILPWTSLDEV  
 EARLIRAGREVEEMIGYELPKPRIIGIMLEVPVMVPLARVDFISVGTSLDLDQ  
 YILAVDRNTRVANIYDSLHPAMRLALAMIRABRIHIGIDLICGEMAGDPMCVAILI

Query Match 95.9%; Score 1374.8; DB 1; Length 292906;

Beet Local Similarity 97.4%; Pred. No. 6.9e-311; Matches 1397; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

1 ATGAAAGTAAAGCTGCGCAGAGTTGAAGTGCAGAGAGTGTGTGTGATGTATG 60  
 260584 ATGAAAGTAAAGCTGCGCAGAGTTGAAGTGCAGAGAGTGTGTGTGATGTATG 260525  
 61 CTGATCGTTACTGAGTACGCGCCCAACAGTGTATCTGCGCGAAGCGCGTCCGCTG 120  
 260524 CTGATCGTTACTGAGTACGCGCCCAACAGTGTATCTGCGCGAAGCGCGTCCGCTG 260465  
 121 GTTAAAGTAAATCAATGGAAGAAAGTCCGCGCGCGCGCTAACGTGGCCATGAATATC 180  
 260464 GTTAAAGTAAATCAATGGAAGAAAGTCCGCGCGCGCGCTAACGTGGCCATGAATATC 260405  
 181 GCTTCTCTCGGTGTAATGACGCGCTGGTGGGTTGACGGGCATTGACGATGACGCGC 240  
 260404 GCTTCTCTCGGTGTAATGACGCGCTGGTGGGTTGACGGGCATTGACGATGACGCGC 260345  
 241 GCGCTGAATTAATCTCTGCGCGACGTCACAGTCAATGCGACTTCTGTTCTGTAACGAC 300  
 260344 GCGCTGAATTAATCTCTGCGCGACGTCACAGTCAATGCGACTTCTGTTCTGTAACGAC 260285  
 301 CATCGACCATTAACAAATTACGGGTACTTCCGCGAACCAACAGCTGATCCGCTGGAT 360  
 260284 CATCGACCATTAACAAATTACGGGTACTTCCGCGAACCAACAGCTGATCCGCTGGAT 260225  
 361 TTGTAAGAAGGTTTGAAGTGTGATCCGACGCGCTGACGACGCGATTATCAGGCG 420  
 260224 TTGTAAGAAGGTTTGAAGTGTGATCCGACGCGCTGACGACGCGATTATCAGGCG 260165  
 421 CTGAGTTGATTTGGCGCGCTGTGCTTTCTGACTACGCCAAAGTGCGCTGGCAAGCGTA 480  
 260164 CTGAGTTGATTTGGCGCGCTGTGCTTTCTGACTACGCCAAAGTGCGCTGGCAAGCGTA 260105  
 481 CAGAGAGATTCGAATCTGGCGCGTAAAGCGGTGTTCGGGTGCGATTGATTCGAAGAGT 540  
 260104 CAGAGAGATTCGAATCTGGCGCGTAAAGCGGTGTTCGGGTGCGATTGATTCGAAGAGT 260045  
 541 ACCGATTTTGAAGCGCTACCGCGCGCTACGCTGTAAACGCCGAATCTCTCGGAATTTGAA 600  
 260044 ACCGATTTTGAAGCGCTACCGCGCGCTACGCTGTAAACGCCGAATCTCTCGGAATTTGAA 259985  
 601 GCTGTGTGCTGTAATGTAAGCCGAGAGAGATTTGTAAGCGCGCGATGAACTGATT 660

259984 GCTGTGTGCTGTAATGTAAGCCGAGAGAGATTTGTAAGCGCGCGATGAACTGATT 259925  
 661 GCCGATTAAGACATCTCGGCTCTGTTAGTACCCGTTCCGACAGGGTATGTGCTGCTG 720  
 259924 GCCGATTAAGACATCTCGGCTCTGTTAGTACCCGTTCCGACAGGGTATGTGCTGCTG 259865  
 721 CAACCGGGTAAAGCGCGCTGATATGCAACCCAGCGAGAGTGTATGACGTTACC 780  
 259864 CAACCGGGTAAAGCGCGCTGATATGCAACCCAGCGAGAGTGTATGACGTTACC 259805  
 781 GGTGCGGCGACACCGGTATTTGCGCTCTGCGCGCAACGCTGACGCGGTATTCGCTG 840  
 259804 GGTGCGGCGACACCGGTATTTGCGCTCTGCGCGCAACGCTGACGCGGTATTCGCTG 259745  
 841 GAAGAAGCCGCTCTTGTGCAATGCGCGCGCTGCGGTGTGTGCGCAACTGGAAC 900  
 259744 GAAGAAGCCGCTCTTGTGCAATGCGCGCGCTGCGGTGTGTGCGCAACTGGAAC 259685  
 901 TCACAGGTTTCGCGCTGCAATGCGCGCTGGAATAATGCTGTACGTGACATACAGGCTTT 960  
 259684 TCACAGGTTTCGCGCTGCAATGCGCGCTGGAATAATGCTGTACGTGACATACAGGCTTT 259625  
 961 GCGGTATGACCGGAAGACATGTAAGCTGCGCGCTGACCGCGCTTAAACGTGTGAA 1020  
 259624 GCGGTATGACCGGAAGACATGTAAGCTGCGCGCTGACCGCGCTTAAACGTGTGAA 259565  
 1021 AAGTGTGATGACCAACGCGTCTTGTGACATCTGTGACCGCGCGCTCTTATCTG 1080  
 259564 AAGTGTGATGACCAACGCGTCTTGTGACATCTGTGACCGCGCGCTCTTATCTG 259505  
 1081 GCAATGCGCGCAAGCTGGGTGACCGCTTGAATTTGTGCGCTCAACAGCATGCTTCACC 1140  
 259504 GCAATGCGCGCAAGCTGGGTGACCGCTTGAATTTGTGCGCTCAACAGCATGCTTCACC 259445  
 1141 AAACGCTGAAGAGGATTTCCCGCGCGGTAAACCCACTGCAACAGCTATGATTGCTG 1200  
 259444 AAACGCTGAAGAGGATTTCCCGCGCGGTAAACCCACTGCAACAGCTATGATTGCTG 259385  
 1201 GCGCACTGGAACGCGTGAAGTGTGCTGTTGAAGAGACAGCGCGAGCGCTTG 1260  
 259384 GCGCACTGGAACGCGTGAAGTGTGCTGTTGAAGAGACAGCGCGAGCGCTTG 259325  
 1261 ATGCGCGGATCTTGCAGATCTGCTGTGTAAGAGCGCGCGACTTAACCAAGAGATT 1320  
 259324 ATGCGCGGATCTTGCAGATCTGCTGTGTAAGAGCGCGCGACTTAACCAAGAGATT 259265  
 1321 GCGCGAGTAAAGAGTCTGGGCGCAACGCTGGGAGAGTGTGGTCTCACTTTGAAGAC 1380  
 259264 GCGCGAGTAAAGAGTCTGGGCGCAACGCTGGGAGAGTGTGGTCTCACTTTGAAGAC 259205  
 1381 GGTGTCTGACGACCAACATCATCAAGAGATCCACAGAGTAAAGGCTAA 1434  
 259204 GGTGTCTGACGACCAACATCATCAAGAGATCCACAGAGTAAAGGCTAA 259151

RESULT 13  
 AB005674\_31/c  
 WPCOMMENT  
 Sequence split into 46 fragments LOCUS AB005674 Accession AB005674

Fragment Name	Begin	End
AB005674_00	1	11000
AB005674_01	10001	21000
AB005674_02	20001	31000
AB005674_03	30001	41000
AB005674_04	40001	51000
AB005674_05	50001	61000
AB005674_06	60001	71000
AB005674_07	70001	81000
AB005674_08	80001	91000
AB005674_09	90001	101000
AB005674_10	100001	111000
AB005674_11	110001	121000
AB005674_12	120001	131000







Details of *S. typhi* sequencing at the Sanger Centre are available on the World Wide Web. ([http://www.sanger.ac.uk/Projects/S\\_typhi/](http://www.sanger.ac.uk/Projects/S_typhi/)).

## FEATURES

## source

Location/Qualifiers

1. 258050

/organism="Salmonella enterica subsp. enterica serovar Typhi"

/mol\_type="genomic DNA"

/strain="CT18"

/db\_xref="taxon:90370"

69. 1544

/gene="tolC"

/note="synonym: STY3364"

69. 1544

/gene="tolC"

/note="Similar to Salmonella enteritidis outer membrane protein TolC precursor tolC SW:tolC\_SALBN (Q54001) (491 aa) fasta scores: E(): 0, 99.04 id in 491 aa, and to Escherichia coli outer membrane protein tolC precursor tolC or mcb or mka or refi SW:tolC\_ECOLI (P02930) (495 aa) fasta scores: E(): 0, 89.74 id in 495 aa Orthologue of E. coli tolC (TOLC\_ECOLI); Fasta hit to TOLC\_ECOLI (495 aa), 90% identity in 495 aa overlap"

/codon\_start=1

/transl\_table=11

/product="outer membrane protein TolC precursor"

/protein\_id="CAD07712.1"

/db\_xref="GI:16504264"

/db\_xref="GOA:Q823N8"

/db\_xref="InterPro:IPR003423"

/db\_xref="InterPro:IPR010130"

/db\_xref="UniProt/TREMBL:Q823N8"

/translation="MOKKLLPILIGSLGFSSTLGAENLMQYQARLSNPELRKSAADDAPEKINERASPLPOLGADYTSNGRDNGINSNETSASLDLTPTLPM SKRGLTLQERAGIDVTYQTDQTLILNTANAYFVLAIDVLSTYQKAIYQO LDTQRFNVLVAITDVQNAQAQYDVLNAYFARNLDNAYELRQVGNYPBLA SLNVEHFKTQPKPKNALKEAENRNLSLOARLSODLAROIRAOQDGLPTLNLA STGISDTSVSGSKTNSAODSDNNGKIGLNSLPYQGMNVSQYKQAOYFVNAS EQLSARISVYQTRSPFNINASTISINAKQAVYSAQSSLDAMEAGYVGRITTYD VLDATTLTDKQQLANARYTTLNQINIRYALGTINRQDLALNSTLGRIPTPSPS VADPTPEQDAADQYNHSAAPAVQPTAANSNNGNPFRR"

1757. 2428

/gene="STY3365"

1757. 2428

/gene="STY3365"

/note="Orthologue of E. coli ygiB (YGI\_B\_ECOLI); Fasta hit to YGI\_B\_ECOLI (234 aa), 97% identity in 223 aa overlap. Contains a possible N-terminal signal sequence."

/codon\_start=1

/transl\_table=11

/product="possible lipoprotein"

/protein\_id="CAD07713.1"

/db\_xref="GI:16504265"

/db\_xref="UniProt/TREMBL:Q8XGZ1"

/translation="MKRTKSIHHSFPRKSAHLLTPVALAVTAVFMLACGKSDRTV SIYONADDCSAAANGKSAECTTAYNNALKEARERAPKATAREDCVAERGGCCQAPR OAGAPENQAOQOQSSGFMWPLAGTWMGLKMGFAQOQPLFSSKNRASPAYGKT TDAAGKNYGAAGPRTMTVPTMAPAPATTYTTVTRGFGESVAKOSTMORSAAGIST RSMGG"

1835. 1867

/gene="STY3365"

/note="PS00013 Prokaryotic membrane lipoprotein lipid attachment site"

2434. 3597

/gene="STY3366"

2434. 3597

/gene="STY3366"

/note="Fasta hit to YJFC\_ECOLI (387 aa), 50% identity in 392 aa overlap Orthologue of E. coli ygiC (YGI\_C\_ECOLI); Fasta hit to YGI\_C\_ECOLI (386 aa), 94% identity in 386 aa overlap"

/codon\_start=1

/transl\_table=11

gene

CDS

/product="conserved hypothetical protein"

/protein\_id="CAD07714.1"

/db\_xref="GI:16504266"

/db\_xref="InterPro:IPR005494"

/db\_xref="UniProt/TREMBL:Q8XRP6"

/translation="MERVSIITERPDRDRALEGFNFHTMYGEPYCEADAYKTLAQ VETLDPYALHOMKLVKVRVVASDELMTKPRIPGHTGFWYQSQWQOOPSRLD LMDGIGEPKLENNADPTSLYEAFQPMIMEDQINGNLPEGSDFNSLOEKLE RPALEBOYGFOLHITCCRDYVEDSGTQYLODCAAEALNTEPYIDIGERGO FTDLOQVIANLFLYMERFLKEMSTKEDGAVWLEPAKSTISNRALLPLNEM FPDHNLPAVFAEDHPMDKVVPRITSRGANVSIENGTTISVEGPGEEEMT VQQPYLPKRGSYTLIGSWLINDQAGIGIREDRALITODLSRFYPHLEVGC"

complement (3659. 4489)

/gene="STY3367"

complement (3659. 4489)

/gene="STY3367"

/note="Orthologue of E. coli ygiD (YGI\_D\_ECOLI); Fasta hit to YGI\_D\_ECOLI (271 aa), 87% identity in 266 aa overlap"

/codon\_start=1

/transl\_table=11

/product="conserved hypothetical protein"

/protein\_id="CAD07715.1"

/db\_xref="GI:16504267"

/db\_xref="GOA:Q823N7"

/db\_xref="InterPro:IPR004183"

/db\_xref="UniProt/TREMBL:Q823N7"

/translation="NMYLTKCLSCSKNINSLTGMPLFLGHSPPAVVLDNDYTRAR RIGBALPRQALVVSAMHYTRGTGTVAHERPOTLHDFGFPALYDMHIFAGSPAL AGRVLELAPVPAVDKEMAGPDHSGMWGLIKKYPNADIPMQLSYDSTKPAWHPEM GRLATLRDGEVNLVAGNVVNLRTVRMHGDNIPYPMAASFNDPVKANTWQGPYEO HPLVNTLQHBGALSNPTPEHFLPLLYLVGAMGKEPITIPVDIGEMGISMSLVQV G"

4588. 5361

/gene="STY3368"

4588. 5361

/gene="STY3368"

/note="Orthologue of E. coli ygiE (YGI\_E\_ECOLI); Fasta hit to YGI\_E\_ECOLI (257 aa), 93% identity in 257 aa overlap. Contains multiple possible membrane spanning hydrophobic domains"

/codon\_start=1

/transl\_table=11

/product="putative membrane protein"

/protein\_id="CAD07716.1"

/db\_xref="GI:16504268"

/db\_xref="GOA:P67471"

/db\_xref="InterPro:IPR003689"

/db\_xref="UniProt/Swiss-Prot:P67471"

/translation="MSVPLILTLAAGATFIGLGVPLDILPAHPODILVOKROOPL LILSLMEMPALDTRGMSPVGYGMFTIGLGVPLDILPAHPODILVOKROOPL GSIKRTAILTLTGISLHNPBEGIATVYTAASSNLELFGIALAVALLNITRGLAVAPV YATGSKRTAIFWAGISGMAELIGVLAALLIGSLVSPVMAIMAAVAGIMVALSV D EIMPLAKRIDPNNNPISYGVLCGMSIMGLSLVILQITIGIC"

6004. 7800

/gene="STY3370"

/note="synonym: asst"

6004. 7800

/gene="STY3370"

/note="Similar to Klebsiella pneumoniae arylsulfate sulfotransferase asst TR:P97036 (EMBL:U32616) (598 aa) fasta scores: E(): 0, 86.84 id in 598 aa"

/codon\_start=1

/transl\_table=11

/product="probable arylsulfate sulfotransferase"

/protein\_id="CAD07717.1"

/db\_xref="GI:16504269"

/db\_xref="GOA:Q823N6"

/db\_xref="InterPro:IPR010362"

/db\_xref="UniProt/TREMBL:Q823N6"

/translation="MPDQYRKTIAGAVALLTGLTAASPAAFGPAQPAQKIGAVV DPAGNPLTALVELDHIISDVYVYVAGGEGVAVPYVYVAGGKSLSTYDGIPLFGLYQ KPAIVTVEYKNGKMDKYDVVQTSALVNHVMDNKSISDLQOTKVIKVAPEGDELY LVTHTFTQGAFFHNGEKDKNAGIILDGAPGALPFDIAPYTFVVDIGEYRWMD



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2006, 20:04:03 ; Search time 550 Seconds  
(without alignments)  
6079.393 Million cell updates/sec

Title: US-10-520-820-28

Perfect score: 1434

Sequence: 1 atgaagaatgaactgcgcaga.....aacagataaataaagctaa 1434

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 8023312 seqs, 1165852854 residues 16046624

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications NA New:

- 1: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1434	100.0	1434	8 US-10-520-820-28	Sequence 28, App1
2	259.4	18.1	377	9 US-11-116-881A-1812	Sequence 1812, Ap
3	253.4	17.7	960	8 US-10-467-657-3253	Sequence 3253, Ap
4	108.2	7.5	507	8 US-10-467-657-547	Sequence 547, App
5	64.6	4.5	510	12 US-11-098-686-8919	Sequence 8919, Ap
6	64.6	4.5	1457619	12 US-11-098-686-8739	Sequence 8739, Ap
7	62.4	4.4	1038	12 US-11-098-686-9716	Sequence 9716, Ap
8	62.4	4.4	1457619	12 US-11-098-686-8739	Sequence 8739, Ap
9	54.6	3.8	200	12 US-11-098-686-6509	Sequence 6509, Ap
10	43.2	3.0	3477	8 US-10-858-730-441	Sequence 141, App
11	37.2	2.6	1140	8 US-11-096-568A-31210	Sequence 31210, A
12	37.2	2.6	2253	9 US-11-096-568A-31087	Sequence 31087, A
13	37	2.6	88421	12 US-11-205-109-1	Sequence 1, App1
14	36.6	2.6	699	8 US-10-467-657-7053	Sequence 7053, App
15	36.6	2.6	3513	8 US-10-467-657-7841	Sequence 7841, Ap
16	36.4	2.5	1349	9 US-11-096-568A-26296	Sequence 26296, A
17	34.8	2.4	200	12 US-11-098-686-1545	Sequence 1545, Ap
18	34.6	2.4	1412	9 US-11-096-568A-10264	Sequence 10264, A
19	34	2.4	789	8 US-10-467-657-4617	Sequence 4617, Ap
20	33.8	2.4	1339	9 US-11-096-568A-6279	Sequence 6279, Ap

C	21	33.6	2.3	200	12	US-11-098-686-1544	Sequence 1544, Ap
C	22	33.4	2.3	200	12	US-11-098-686-6513	Sequence 6513, Ap
C	23	33.4	2.3	1632	9	US-11-096-568A-12879	Sequence 12879, A
C	24	33.4	2.3	1957	8	US-10-750-185-42527	Sequence 42527, A
C	25	33.4	2.3	1957	8	US-10-750-623-42527	Sequence 42527, A
C	26	33.2	2.3	1515	9	US-11-179-363-4	Sequence 4, App1
C	27	33.2	2.3	18394	8	US-10-995-561-13367	Sequence 13367, A
C	28	33.2	2.3	400	8	US-10-953-392-1	Sequence 17, App1
C	29	33	2.3	400	8	US-10-954-147-17	Sequence 17, App1
C	30	33	2.3	2196	12	US-11-052-554A-539	Sequence 539, App
C	31	32.8	2.3	813	8	US-10-467-657-2265	Sequence 2265, Ap
C	32	32.8	2.3	1551	9	US-11-096-568A-12361	Sequence 12361, A
C	33	32.8	2.3	2465	9	US-11-072-512-1278	Sequence 1278, Ap
C	34	32.6	2.3	678	8	US-11-187-622-14	Sequence 5, App1
C	35	32.6	2.3	879	8	US-10-965-972-5	Sequence 23715, A
C	36	32.6	2.3	1161	9	US-11-096-568A-23715	Sequence 20738, A
C	37	32.6	2.3	1165	9	US-11-096-568A-20738	Sequence 20738, A
C	38	32.6	2.3	5071	12	US-11-090-617-563	Sequence 563, App
C	39	32.6	2.3	76427	12	US-11-124-367A-5041	Sequence 5041, Ap
C	40	32.6	2.3	415117	8	US-10-995-561-13374	Sequence 13374, A
C	41	32.4	2.3	432	8	US-10-467-657-3275	Sequence 3275, Ap
C	42	32.4	2.3	438	8	US-10-953-392-14	Sequence 14, App1
C	43	32.4	2.3	438	8	US-10-954-147-14	Sequence 14, App1
C	44	32.4	2.3	1213	7	US-10-537-897-18	Sequence 18, App1
C	45	32.4	2.3	1213	7	US-10-537-897-36	Sequence 36, App1

#### ALIGNMENTS

RESULT 1	US-10-520-820-28	Application US/10520820
Sequence 28, Application US/10520820	US-10-520-820-28	US20060003393A1
PUBLICATION NO. US20060003393A1		
GENERAL INFORMATION:		
APPLICANT: MUTABILIS S. A.		
TITLE OF INVENTION: Pathogenicity determinants which can be used as targets for develop		
TITLE OF INVENTION: means for preventing and controlling bacterial infections and/or		
TITLE OF INVENTION: dissemination		
FILE REFERENCE: 1621		
CURRENT APPLICATION NUMBER: US/10/520, 820		
CURRENT FILING DATE: 2005-01-07		
PRIOR APPLICATION NUMBER: PCT/EP2003/008209		
PRIOR FILING DATE: 2003-07-09		
PRIOR APPLICATION NUMBER: FR 0208636		
NUMBER OF SEQ ID NOS: 32		
SOFTWARE: PatentIn version 3.1		
SEQ ID NO 28		
LENGTH: 1434		
TYPE: DNA		
ORGANISM: Escherichia coli		
US-10-520-820-28		
Query Match	100.0%	Score 1434; DB 8; Length 1434;
Best Local Similarity	100.0%	Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;
Matches 1434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	1	ATGAAAGTAAAGCTGTCAGAGTTTAAACGTGACAGAGTATGTGTGTGTATGTATG 60
DB	1	ATGAAAGTAAAGCTGTCAGAGTTTAAACGTGACAGAGTATGTGTGTGTATGTATG 60
QY	61	CTGAGATCTTACCTGTCAGAGTTTAAACGTGACAGAGTATGTGTGTGTATGTATG 120
DB	61	CTGAGATCTTACCTGTCAGAGTTTAAACGTGACAGAGTATGTGTGTGTATGTATG 120
QY	121	GTTAAAGTAAATTCATCAAGAAAGCTGCGGCGCGCGCTTAACTGACATGATATC 180
DB	121	GTTAAAGTAAATTCATCAAGAAAGCTGCGGCGCGCGCTTAACTGACATGATATC 180
QY	181	GCTTCTCTCGGTCTATATCAAGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 240
DB	181	GCTTCTCTCGGTCTATATCAAGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 240

QY	241	CCGCTGAGTAATCTCTGCGCGAGCTGAACGTCAAATGCGACTTCGTTCTGTACCCAGC	300
Db	241	GGCTGAGTAATCTCTGCGCGAGCTGAACGTCAAATGCGACTTCGTTCTGTACCCAGC	300
QY	301	CATCCGACCATTAACCAAAATTAACGGGTACTTTCGCGCAACCAACAGCTGATCCGTGGAT	360
Db	301	CATCCGACCATTAACCAAAATTAACGGGTACTTTCGCGCAACCAACAGCTGATCCGTGGAT	360
QY	361	TTTGAAGAAAGTTTCCGAAGGTGTGATCCGCAAGCCGCTGCAACAGCGGATTTAATCAGCGC	420
Db	361	TTTGAAGAAAGTTTCCGAAGGTGTGATCCGCAAGCCGCTGCAACAGCGGATTTAATCAGCGC	420
QY	421	CTGAGTTGATTTGGCGCGCTGGTGCTTCTGACATACGCGCAAAAGTGCGCTGGCAAGCGTA	480
Db	421	CTGAGTTGATTTGGCGCGCTGGTGCTTCTGACATACGCGCAAAAGTGCGCTGGCAAGCGTA	480
QY	481	CAGCAGATGATCCAACTGGCGCGTAAAGCGGGTGTCCGGTCTGATTTGATCCAAAAGT	540
Db	481	CAGCAGATGATCCAACTGGCGCGTAAAGCGGGTGTCCGGTCTGATTTGATCCAAAAGT	540
QY	541	AACGATTTTGAAGCGGTACCGCGCGCTACCGCTGTTAAGCCGAATCTCTCGAAATTTGAA	600
Db	541	AACGATTTTGAAGCGGTACCGCGCGCTACCGCTGTTAAGCCGAATCTCTCGAAATTTGAA	600
QY	601	GCTGTGTCCGTTAATGTAAAGCCGAAGAGATTTGTGAAGCGCGCATGAACTGATTT	660
Db	601	GCTGTGTCCGTTAATGTAAAGCCGAAGAGATTTGTGAAGCGCGCATGAACTGATTT	660
QY	661	GCCGATTAACGAATCTCTGGCTCTGTATGACCCGTTCCGAACAGGGTATGTCCGTGCTG	720
Db	661	GCCGATTAACGAATCTCTGGCTCTGTATGACCCGTTCCGAACAGGGTATGTCCGTGCTG	720
QY	721	CAACCGGGTTAAAGCGCGCTCATATGCGCAACCCAGCGCAGAAAGTTATGACGTTACC	780
Db	721	CAACCGGGTTAAAGCGCGCTCATATGCGCAACCCAGCGCAGAAAGTTATGACGTTACC	780
QY	781	GGTGCGGGCGACACGGTGATTTGGCGCTCTGGCGCGCAACGCTGGCAAGCGGTTAATTCGCTG	840
Db	781	GGTGCGGGCGACACGGTGATTTGGCGCTCTGGCGCGCAACGCTGGCAACGCGGTTAATTCGCTG	840
QY	841	GAAAGAGCTGCTCTTTTGGCCAAATGCGGCGCGCTGGCGGTGATTCGCGCAACTGGGAACC	900
Db	841	GAAAGAGCTGCTCTTTTGGCCAAATGCGGCGCGCTGGCGGTGATTCGCGCAACTGGGAACC	900
QY	901	TCCACGGTTTCCGCCGATGAGCTGGAAATGTCTGTACGTGACCGTGCAGATTAACAGGCTTT	960
Db	901	TCCACGGTTTCCGCCGATGAGCTGGAAATGTCTGTACGTGACCGTGCAGATTAACAGGCTTT	960
QY	961	GGCGGTATGACCGGAAGAGAACTGAAGCTGGCGCGTGAAGCGCGCGCTTAAACGTGTGAA	1020
Db	961	GGCGGTATGACCGGAAGAGAACTGAAGCTGGCGCGTGAAGCGCGCGCTTAAACGTGTGAA	1020
QY	1021	AAAGGTGTGATGACCAACGAGTGTCTTTGACATCTCGACAGCGCGGCAAGTCTTTATGTG	1080
Db	1021	AAAGGTGTGATGACCAACGAGTGTCTTTGACATCTCGACAGCGCGGCAAGTCTTTATGTG	1080
QY	1081	GCAAAATGCCCGCAGCTGGGTGACCGCTTGATTTGGCTGTCAACAGCGATGCTCCACC	1140
Db	1081	GCAAAATGCCCGCAGCTGGGTGACCGCTTGATTTGGCTGTCAACAGCGATGCTCCACC	1140
QY	1141	AAAAGGCTGAAAAGGGGATTTCCCGCCCGGTTAAACCACTCGAACAGCGATATGTGTGCTG	1200
Db	1141	AAAAGGCTGAAAAGGGGATTTCCCGCCCGGTTAAACCACTCGAACAGCGATATGTGTGCTG	1200
QY	1201	GGCGCACTGGAAGCGGTGACCTGGGTAGTGTCTTTGAAGAGACACGCGCGAAGGCTTGG	1260
Db	1201	GGCGCACTGGAAGCGGTGACCTGGGTAGTGTCTTTGAAGAGACACGCGCGAAGGCTTGG	1260
QY	1261	ATCGCGGGGATCTTCCGACGATTTGCTGTGTGAAGCGCGCGCATATTAACAGAAAGATTT	1320
Db	1261	ATCGCGGGGATCTTCCGACGATTTGCTGTGTGAAGCGCGCGCATATTAACAGAAAGATTT	1320

```

OY      1321  GC CGGAGCTAAAGAAAGCTCTGGCCCAACGGTGGGGAAGTGTGGGCTCACTTTGAAGAC  1380
DB      1321  GCCCGGAGTAAAGAGTCTGGGCCCAACGGTGGGGAAGTGTGGGCTCACTTTGAAGAC  1380
OY      1381  GGTGCTCGACGACCAATCATCAAGAGATCCAAAGAGTAAAAAGGCTAA  1434
DB      1381  GGTGCTCGACGACCAATCATCAAGAGATCCAAAGAGTAAAAAGGCTAA  1434

RESULT 2
US-11-116-881A-1812
; Sequence 1812, Application US/11116881A
; Publication No. US20060041949A1
GENERAL INFORMATION:
APPLICANT: Xu, Dongmei
APPLICANT: Nielsen, Mark T.
TITLE OF INVENTION: Nicotiana Nucleic Acid Molecules and Uses Thereof
FILE REFERENCE: 07678/141014
CURRENT APPLICATION NUMBER: US/11/116, 881A
CURRENT FILING DATE: 2005-04-27
PRIOR APPLICATION NUMBER: 60/665,451
PRIOR FILING DATE: 2005-03-24
PRIOR APPLICATION NUMBER: 60/665,097
PRIOR FILING DATE: 2005-03-24
PRIOR APPLICATION NUMBER: 60/646,764
PRIOR FILING DATE: 2005-01-25
PRIOR APPLICATION NUMBER: 60/607,357
PRIOR FILING DATE: 2004-09-03
PRIOR APPLICATION NUMBER: 60/566,235
PRIOR FILING DATE: 2004-04-29
PRIOR APPLICATION NUMBER: 10/934,944
PRIOR FILING DATE: 2004-09-03
PRIOR APPLICATION NUMBER: 10/943,507
PRIOR FILING DATE: 2004-09-17
PRIOR APPLICATION NUMBER: 60/503,989
PRIOR FILING DATE: 2003-09-18
PRIOR APPLICATION NUMBER: 60/485,368
PRIOR FILING DATE: 2003-07-08
PRIOR APPLICATION NUMBER: 60/418,933
PRIOR FILING DATE: 2002-10-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 2300
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1812
; LENGTH: 377
; TYPE: DNA
; ORGANISM: Nicotiana tabacum
US-11-116-881A-1812

Query Match      18.1%; Score 259.4; DB 9; Length 377;
Best Local Similarity 99.6%; Pred. No. 3,6e-70;
Matches 260; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      1174  CCACTCGAAGCGGTATGATTTGCTGGGCGCACTGGAACGGTGCATG3GTAAGTGTGC  1233
DB      1      CCACTCGAAGCGGTATGATTTGCTGGGCGCACTGGAACGGTGCATG3GTAAGTGTGC  60
OY      1234  TTTGAAGAAGACACAGCCGCGACCGCTTGAATGGCCGGATCTTTGCCAAGATCTGCTGTA  1293
DB      61      TTTGAAGAAGACACAGCCGCGACCGCTTGAATGGCCGGATCTTTGCCAAGATCTGCTGTA  120
OY      1294  GCGCGGCACTTAACACAGAAAGAGATTGGCGGAGTAAAGAGTCTGGGCCAACGGTGGC  1353
DB      121  GCGCGGCACTTAACACAGAAAGAGATTGGCGGAGTAAAGAGTCTGGGCCAACGGTGGC  180
OY      1354  GAAAGTGTGGTGTCTCAACTTTGAAGACGTTGCTCGACGACCAACATCATCAAGAGATC  1413
DB      181  GAAAGTGTGGTGTCTCAACTTTGAAGACGTTGCTCGACGACCAACATCATCAAGAGATC  240
OY      1414  CAACAGATTAAGGCTAA  1434
DB      241  CAACAGATTAAGGCTAA  261

```

```

Qy      682 CTGTTAGAGCAACCGTTCCGAACAGGGTATGTGCTGCTGCAACCCGGGTAAACGCGCGCTG 741
Db      700 GTTTTGTGCAACCCGGAGGAGAGAGGGCATGACCCGTGTCAGGAAAGGCGAA---CCCAATT 756
Qy      742 CATATGCCAACCCAGCGCAGAGAGTATATGACGTTAACCGGTGCGGGCGACACGGTGATT 801
Db      757 TACACAGCCCAACCCGGGCCCAAGAAATTATACGACGTGTCCGGGCGACGGCACACCGTCATT 816
Qy      802 GGCGGCTCGGCGGCGCAACGCTGGCAGCGGGTAAATGTGCTGAAAGAACCTGTTCTTTGGC 861
Db      817 GCCGGAATGGGCTTGGGGGCTGGCGGCAAGCTGCACATATCCGAAAGCCATGTAACTTTGCC 876
Qy      862 AATGCGGCGGCTGCGGCTGTGCTGTGCTGCGCAAACTGGAGAACCTCCACCGGTTTGGCGGATCGAG 921
Db      877 AATATCGGCGCGGGGTTGTGTGTGCGCAAACTGTGTAAGGCGGCTTTGTCTGTTTGACAGAG 936
Qy      922 CTGGAATAATGC 932
Db      937 TTGGTTGAAGC 947

```

```

RESULT 4
US-10-467-657-547
; Sequence 547: Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; PRIOR FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: Seqwinn9, version 1.04
; SEQ ID NO 547
; LENGTH: 507
; TYPE: DNA
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-547

Query Match          7.5%; Score 108.2; DB 8; Length 507;
Query Local Similarity 56.5%; Pred. No. 3e-23; Mismatches 168; Indels 3; Gaps 1
Matches 222; Conservative 0;

Dy      1025 TGTGATGATGACCAACGGTGTCTTTTGACATCTCTGACGCGGGCAGCTCTTATCTGGCA 108
Dy      95  TGTGTTTACCAACGGCTGTTTGACATCTCTCAAGGGGGCAGCTTATCTGGGCG 154

Qy      1085 ATGCCCCGAAAGCTGGGTGACCGCTTGATTGTTGCCGTCAACAGCATGCTTCACCAAC 114
Dy      155  AGGCGCGCTTCGGCAGAGGGCGCGCGTTGGTGCTGGCGTTGAATACGATCTTCGGTGGCG 214

Qy      1145 GGCCTG---AAAGGGGATTCGCCGCCCGGTAACCCACTGTAACAGGTATGATGTCG 120
Dy      215  GTTTGGGCAAGGGGGGGGCGACCGCCCGGTTAATCTTTTGAAGAACCGTCCCGCGTTGGCG 274

Qy      1202 GCGCAGCTGGAAGCGGTGCGACTGCGGTAGTGTGTTTGAAGAGACAGCGCGAGCGCTTGA 126
Dy      275  CCGGCTTGGAAGAGTGTGATTTGGTTACTTGTTGACAGAGATACGCGCGGGGGTTGA 334

Qy      1262 TCGCGCGGATCTTTCAGATCTGCTGTGTGAAGAGCGGCGACTATTAACCAAGAGATTG 1322
Dy      335  TTGAGCGGCTCAACCGCGAGGTGTTGGTCAAGGCGGCGATTGGGTGTTGATTAAGATTG 394

Qy      1322 CCGGGAATTAAGAGTCTGGGCGCAACGCTGGCGAAGTGTGTGCTCACTTTGAAGACG 1388
Dy      395  TCGGTGCGCGGGAACCGCTGGCAGCGCGGCTCAAGTGTTTTCAATCCGTTTCTGCAATC 454

```







Db	67	GTCTTGTCGTTGGTGATATTAATGTAATTAATTATCTTAATGGGTTCAGATCGTAAT	126
Qy	97	TTCCGCGGAAGCCGCGGTGCGCGGTGAATAGTAATACCATCGAAGAACGTCGCGGGCGC	156
Db	127	TCTCTGAGGACCTCTCCCTATGTAATAAATAGAAATGAAGAACAACTCTCTTGGGGGGT	186
Qy	157	GCGGCTAAACGTGCGATGAAATATCGCTTCTCTCGGTGCTAAATGCACGCTGTCGGGTG	216
Db	187	GCAGGAAATGTAGCAAGAGTATGTCGTCTCGGTGGAAGGTAAACGATTAATGGGGCT	246
Qy	217	ACGGGCAATTGACGATGCACGCGCGCGCTGAAATAATCTGCGCGCAACTGAAGTCAAA	276
Db	247	GTAAGACAAACCAAGGTGAGAAAAATTCAAGATCTTTTAAGTACAAGGGAAATTTTA	306
Qy	277	TGGCACTTCGTTCTGTACGACGCAATCCGACATTACAATAATACGGGTACTTTCGCGC	336
Db	307	TCTTCATTTTAACATTTCGCAATCGTCAACAAACAGTAATAAACACGTGATGGCTCAC	366
Qy	337	AACCAACAGCTGATCGGTCTGGAATTTTGAAGAAAGT	372
Db	367	AGACAAAGATGATACGACTTGATCTACATGAGGAAGT	402

```

RESULT 8
US-11-098-686-8739
Sequence 8739, Application US/11098686
Publication No. US20050024696A1
GENERAL INFORMATION:
APPLICANT: Kapur, Vivek and Gebhart, Connie J.
TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
TITLE OF INVENTION: FROM LAWSONIA INTRACELLULARIS AND METHODS OF USING
FILE REFERENCE: 09531-128001
CURRENT APPLICATION NUMBER: US/11/098,686
CURRENT FILING DATE: 2005-04-04
PRIOR APPLICATION NUMBER: PCT/US00/31318
PRIOR FILING DATE: 2003-10-01
PRIOR APPLICATION NUMBER: US 60/416,395
PRIOR FILING DATE: 2002-10-04
NUMBER OF SEQ ID NOS: 11433
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 8739
LENGTH: 1457619
TYPE: DNA
ORGANISM: Lawsonia intracellularis
US-11-098-686-8739

```

	Query Match	4.4%;	Score 62.4;	DB 12;	Length 1457619;
	Best Local Similarity	49.1%;	Pred. No. 6.6e-08;		
	Matches 165;	Conservative	0;	Mismatches 171;	Indels 0;
				Gaps	0;
QY	37	GTGATGGTGGTGGTGGATGATGATCTGGATCGTTACTGGTACGGCCCCACGATCGTATC	96		
Db	1014025	GTTCTTTCGTTGGTGGATATTAATGTTAGTAAATTATCTATTATGGTGTTCAGATCGTATTT	1014084		
QY	97	TCGCCCGGAAGGCCGGTGCCTGGTGTAAAGTGAATACATCGAAGAACGTCCGGCGGC	156		
Db	1014085	TCTCTTGAGGACCTGTCTCCCTATTGTAAAAATAGAGATGAAAAAGCAATCTCTTGGGGGGT	1014144		
QY	157	GGCGCTAACGTCGGGGAATGAATATGGCTCTCTCGGTGCTAATGCAACGCTGTCGGGGTGG	216		
Db	1014145	GCAGGGAATGTAGCAAGGATATTGCTGCTCTGGGTGGAAGGTACATTAATTTGGGGCTT	1014204		
QY	217	ACGGGCAATTGACGATGACAGCCGCGCGCTGAGTAATCTCTGCGCGCACTGCAACGTCAAA	276		
Db	1014205	GTAGAGCAAGACCAAAAGTGAAGAAAAAATTCAAGTCTTTTAAGTACAAAGGGAATTTTA	1014264		
QY	277	TGCGACTTCGTTTCTGTACCGACGCGATCCGACATACCAAAATTACGGGTACTTTCCCGC	336		
Db	1014265	TCTTCAATTATAACATTTTGCAAACTGTCMAACAACAGTAAAAACAACGTCGTATGGCTCAC	1014324		
QY	337	AACCAACGCTGATCCCTGTCGTGGATTTTGAAGAAAGT	372		
Db	1014325	AGACAAACGATGATACGACTTGATGATCAGAGAAAGT	1014360		

```

RESULT 9
US-11-098-686-6509
; Sequence 6509, Application US/11098686
; Publication NO. US20060024656A1
; GENERAL INFORMATION:
; APPLICANT: Kapur, Vivek and Gebhart, Connie J.
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
; TITLE OF INVENTION: FROM LAWSONIA INTRACELLULARIS AND METHODS OF USING
; FILE REFERENCE: 09531-128001
; CURRENT APPLICATION NUMBER: US/11/098,686
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: PCT/US03/31318
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: US 60/416,395
; PRIOR FILING DATE: 2002-10-04
; NUMBER OF SEQ ID NOS: 1433
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6509
; LENGTH: 200
; TYPE: DNA
; ORGANISM: Lawsonia intracellularis
US-11-098-686-6509

```

Query Match	3.8%	Score 54.6	DB 12	Length 200
Best Local Similarity	58.2%	Pred. No. 9.3e-07		
Matches	96	Conservative	0	Mismatches 69; Indels 0; Gaps 0

  

QY	37	GTGATGATGTTGTGATGTGATGTGATGCTGATGCTTACGTGACGCCGCCACCAAGTCGTATC	96
Db	25	GTTCTTGTCGTTGGTGATATATGTTAATATTTATCTTATGTGTGTTCCAGATCGTATC	84
QY	97	TGCGCGGAAGCCCGGTGCGCGGTGTTAAAGTAATACATCGAAGAAAGTCCGGGCGGC	156
Db	85	TCTCCGAGGCACTGTCCTTATGTAAAAATAGAGATGAAAGCAATCTTGGGGGT	144
QY	157	GGGGCTAAGTGGGCAATGAATATGCGCTCTCGGTGCTATGCA	201
Db	145	GCAAGGAAATGTACAGAGATATCTCTCTCGGCTGGAAGGTA	189

```

RESULT 10
US-10-858-730-141
Sequence 141, Application US/10858730
Publication No. US2005025566A1
GENERAL INFORMATION:
APPLICANT: Bailey, Richard B.
APPLICANT: Blomquist, Paul
APPLICANT: Doten, Reed
APPLICANT: Driggers, Edward M.
APPLICANT: Madden, Kevin T.
APPLICANT: O'Leary, Jessica
APPLICANT: O'Toole, George
APPLICANT: Trueheart, Joshua
APPLICANT: Walbridge, Michael J.
APPLICANT: Yorgey, Peter S.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR AMINO ACID
TITLE OF INVENTION: PRODUCTION
FILE REFERENCE: 14184-030001
CURRENT APPLICATION NUMBER: US/10/858, 730
CURRENT FILING DATE: 2004-06-01
PRIOR APPLICATION NUMBER: US 60/475, 000
PRIOR FILING DATE: 2003-05-30
PRIOR APPLICATION NUMBER: US 60/551, 860
PRIOR FILING DATE: 2004-03-10
NUMBER OF SEQ ID NOS: 364
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 141
LENGTH: 3477
TYPE: DNA
ORGANISM: Thermobifida fusca
US-10-858-730-141

```

Query Match 3.0%; Score 43.2; DB 8; Length 3477;  
Best Local Similarity 46.2%; Pred. No. 0.0082;  
Matches 144; Conservative 0; Mismatches 168; Indels 0; Gaps 0;

QY 1053 CTTGACGCGCGGACGCTCTTATCTGGCAATGCCCGACAGTGGGTGACCGCTTGAT 1112  
DB 1131 CATCTCGACTGCTGCGCACTATGTGGAGCAGCGGGGTGGCGATATGCGGAGCT 1190  
QY 1113 TGTTCGGCTCAACAGCGATGCTTCCACCAACCGCTGAAGGGATTCGCCCGCGTAAA 1172  
DB 1191 GGGTCCCGGTGCGACCGGCTCCACGCTGCGCTGATGACTCCAGAAAGTAC 1250  
QY 1173 CCCACTCGAACAGGCTATGATTTGTCTGGCGCACTGGAACCGGTGACTGGGTATGTC 1232  
DB 1251 GGTACTGGAAGCTGAGACTGAGATGCTGGGCGGCGCGCTGCTCAACTCGTCACTA 1310  
QY 1233 GTTTGAAGAGCAGCAGCGCGACGCTTGATGCGCGGATCTTGGCAGATCTGTGTGAA 1292  
DB 1311 CGAGGACGGCGACCGCCCTGACTCCCGGTTGCGCAAGGTGCGCGCTGGCGGTGAGCA 1370  
QY 1293 AGGCGGCGACTATTAACGAGAGATTTCCGGGAGTAAAGATCTGGGCGCAACGGTGG 1352  
DB 1371 CGGGGCGGCGCTCATGTGGCGCTGACCATGACAGCAGGGGCGAGCGCGGAGACG 1430  
QY 1353 CGAAGTGTGGT 1364  
DB 1431 GAAAGTGAAGT 1442

## RESULT 11

US-11-096-568A-31210  
; Sequence 31210, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 31210  
; LENGTH: 1140  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(1140)  
; OTHER INFORMATION: Ceres Seq. ID no. 13579027  
US-11-096-568A-31210

Query Match 2.6%; Score 37.2; DB 9; Length 1140;  
Best Local Similarity 49.5%; Pred. No. 0.41;  
Matches 96; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 679 GCTCTGTTAGTACCCGTTCCGAAACAGGGTATGCTGCTGCAACCGGTTAAAGCCCG 738  
DB 862 GTTCTAGTAAACTGGGTCCAAAGATCTGCATTTATACAAAGGAAAAACCAATC 921  
QY 739 CTGCATATGCCAACCAAGCCAGAGAGTATGACGTTACCGGTGGCGGACACGGTG 798  
DB 922 CAGCAGTCTATCATACCGGCTGCAACAAGTGTGTAATAAGAGGCTGGGATATCTTC 981  
QY 799 ATTGGCGTCTGGCGGCAAGCGTGGCAGCGGTTAATTGCTGGAAGAACCTGCTTCTT 858  
DB 982 ACTCAGCATTTGACAGTGTGATGTAGAGGCAAGTCCCATGAGAAATGCTTGAGATTT 1041  
QY 859 GCCAATGGCGGCG 872  
DB 1042 GCTGCTGCACTGC 1055

## RESULT 12

US-11-096-568A-31087  
; Sequence 31087, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 31087  
; LENGTH: 2253  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(2253)  
; OTHER INFORMATION: Ceres Seq. ID no. 4987027  
US-11-096-568A-31087

Query Match 2.6%; Score 37.2; DB 9; Length 2253;  
Best Local Similarity 49.5%; Pred. No. 0.52;  
Matches 96; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 679 GCTCTGTTAGTACCCGTTCCGAAACAGGGTATGCTGCTGCAACCGGTTAAAGCCCG 738  
DB 1975 GTTCTAGTAAACTGGGTCCAAAGATCTGCATTTATACAAAGGAAAAACCAATC 2034  
QY 739 CTGCATATGCCAACCAAGCGCAGAGAGTATGACGTTACCGGTGGCGGACACGGTG 798  
DB 2035 CACGAGTCTATCATACCGGCTGCAACAAGTGTGATATACAGAGCTGGGATATCTT 2094  
QY 799 ATTGGCGTCTGGCGGCAACGCTGCGAGCGGTTAATTGCTGGAAGAACCTGCTTCTT 858  
DB 2095 ACTCAGCATTTGACAGTGTGATGTAGAGGCAAGTCCCATGAGAAATGCTTGAGATTT 2154  
QY 859 GCCAATGGCGGCG 872  
DB 2155 GCTGCTGCACTGC 2168

## RESULT 13

US-11-205-109-1  
; Sequence 1, Application US/11205109  
; Publication No. US20050287641A1  
; GENERAL INFORMATION:  
; APPLICANT: Farnet, Chris  
; APPLICANT: Zazopoulos, Emmanuel  
; TITLE OF INVENTION: GENE CLUSTER FOR RAMOPLANIN BIOSYNTHESIS  
; FILE REFERENCE: 3002-205  
; CURRENT APPLICATION NUMBER: US/11/205,109  
; CURRENT FILING DATE: 2005-08-17  
; PRIOR APPLICATION NUMBER: US/09/976,059  
; PRIOR FILING DATE: 2001-10-15  
; PRIOR APPLICATION NUMBER: US 60/239,924  
; PRIOR FILING DATE: 2000-10-13  
; NUMBER OF SEQ ID NOS: 46  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 88421  
; TYPE: DNA  
; ORGANISM: Actinoplanes sp.  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (2077)..(3078)  
; OTHER INFORMATION: ORF 1; positive strandedness  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (3118)..(4032)  
; OTHER INFORMATION: ORF 2; positive strandedness

```

NAME/KEY: misc feature
LOCATION: (74216)..(73563)
OTHER INFORMATION: ORF 21; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (75424)..(74213)
OTHER INFORMATION: ORF 22; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (75535)..(76464)
OTHER INFORMATION: ORF 23; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (78110)..(76449)
OTHER INFORMATION: ORF 24; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (79864)..(78107)
OTHER INFORMATION: ORF 25; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (81624)..(79861)
OTHER INFORMATION: ORF 26; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (81909)..(81682)
OTHER INFORMATION: ORF 27; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (82346)..(82062)
OTHER INFORMATION: ORF 28; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (82587)..(84446)
OTHER INFORMATION: ORF 29; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (84481)..(85548)
OTHER INFORMATION: ORF 30; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (85556)..(86845)
OTHER INFORMATION: ORF 31; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (87372)..(86803)
OTHER INFORMATION: ORF 32; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (87494)..(88420)
OTHER INFORMATION: ORF 33; positive strandedness, N-terminus only
US-11-205-109-1

Query Match      2.6%; Score 37; DB 12; Length 88421;
Best Local Similarity 43.1%; Pred. No. 2; Mismatches 235; Indels 0; Gaps 0;
Matches 178; Conservative

Oy CGAAGCGCCGCGGCGCGGTGAATTCATCAAGAAGACGTCCGGCGGCGCG 160
Db CGGTCCGCGGTGGCGGCGAGGTGTCTGTCTCATGCAGCTTACCGTCCGCG 63551
Oy CTAACGTGGCAGATAATTCCTTCCTCGTGCTAATCAAGCCTGATCGGGTTGACGG 220
Db ATGAGCTGGCGGAGACTGTGTGGCGGAGTGTGGAGTGTGAGTGTGAGTGTGG 63611
Oy GCATTGACATGACGCGCGCGCTGTGAAATCTTGTGCGCACACGTCAATGTGG 280
Db CCTCGAACGCTCCGCGGACGTGTACCCGCTTCTGCCCGTGGAAAGCGCGGCTG 63671
Oy ACTTCGTTTTGTATACGACGACATCCGACCATTCACAATTAAGGGGTACTTTCGCGAAC 340
Db TCTTCGTTCCGGGTGACTTCTCCTGACCAGACGCGGTATCGATGGGTGATCGCGGACA 63731

```

[illegible]

```

RESULT 14
US-10-467-657-7053
; Sequence 7053, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqMan99, version 1.04
; SEQ ID NO 7053
; LENGTH: 699
; TYPE: DNA
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-7053

```

	Query Match	Similarity	2.6%	Score 36.6	DB 8	Length 699
	Match	Local	Similarity	47.6%	Pred. No. 0.54	
	Matches 108	Conservative	0	Mismatches 119	Indels 0	Gaps 0
Qy	1061	CCGGGCAAGTCTTATCTGCAAAATGCCGCAAGTGGGTGACCGCTTGAATTGTGGCG				1120
Db	440	CCGGTCAGGTGTGCAATCGGTGTAAGCGCGCATTTGCCGAGATGGCAGATATCTTTCG				499
Qy	1121	TCAACAGGATAGCTCCACCAAAAGGCGTGGAAAGGGGATTCGCCGCCGTAAACCACTCG				1180
Db	500	TGTATGTCCGACCAACGCCCAACCGACCGGGGTTACTATATTTATGTAAAGAAAGCG				559
Qy	1181	AACAGCGTATGATTGTGCTGGGCGCACTGAAAGCGGTGACTGGGTAGTGTGTTTAAAG				1240
Db	560	ATGTGCGGAACTCGATTTGAGCGGTGACGAAGCGTTGAATATATGTATTTTCGCTGGGTA				619
Qy	1241	AGGACACCGCCGACAGCTTGAATTCGCCGGGATCTTGCCGACATCTGTG				1287
Db	620	TGTGTATCCCTGACGACCTGCGCCGTCAAAACATTTGGCAGGACCTTATG				666

RESULT 15  
US-10-467-657-7841  
Sequence 7841, Application US/10467657  
Publication No. US20050260581A1  
GENERAL INFORMATION:  
APPLICANT: CHIRON SPA  
APPLICANT: FONTANA Maria Rita  
APPLICANT: PIZZA Mariagrazia  
APPLICANT: MASTIGNANI Vega  
TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS  
FILE REFERENCE:  
CURRENT APPLICATION NUMBER: US/10/467,657  
CURRENT FILING DATE: 2003-08-11  
PRIORITY APPLICATION NUMBER: GB-0103424.8

```

; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWing99, version 1.04
; SEQ ID NO 7841
; LENGTH: 3513
; TYPE: DNA
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-7841

```

Query Match	2.6%	Score 36.6	DB 8	Length 3513
Best Local Similarity	45.6%	Pred. No. 0.91		
Matches 129	Conservative	0	Mismatches 154	Indels 0
			Gaps 0	
Qy	553	CGCTACCGGGCGCTACGCTGTTAACGCCGATCTCTCGAATTGAAAGCTGTTCGGT	612	
Db	1588	CGCGATGTGCGCGAAGCGCAACCGATTAAAAACCTGCGACACACGGGATGTGTGGT	1644	
Qy	613	AAATGTAAGACCGAAGAAGATGTTTGAAGCGCGGATGAATCTGATTCGCGATTACGA	672	
Db	1648	ACGGTCAGCTTTGGCGAATGCCGCGCTTGCCCAABAAGGGTGTGGTGAAGCGGTGCCG	1707	
Qy	673	CTCTCGCTCTGTTAGTAGACCCGTTCCGACAGGGTATGTCTGCTGTCGAACCGGGTAA	732	
Db	1708	TTCCCCGAATAGATGTGCGACACTGCGCGCGAAGCGCGCGCTCTGTCGCGCTTCCGC	1767	
Qy	733	GCGCGCGTGCATATGATGCCAACCCAGCGCAGAAATGTATATGACGTTAACCGTGGCGCGAC	792	
Db	1768	GACCTGTCTCGAACAGCACACCCCGCGCTGTATGATGCTTGTCCGTGCGGAAGCGGCAAA	1827	
Qy	793	ACGGTATGTCGCTCTGCGCGCAACGCTGACACGGGTATTT	835	
Db	1828	ACGCTGAACAACGCCATTTGCCGAATGTCCGAAGCGCAACCGCTCGATT	1870	

Search completed: March 18, 2006, 20:19:09  
Job time : 554 secs